

© 2010 Romina Nikoukar

NEAR-OPTIMAL INVERSION OF INCOHERENT SCATTER RADAR
MEASUREMENTS: CODING SCHEMES, PROCESSING TECHNIQUES, AND
EXPERIMENTS

BY

ROMINA NIKOUKAR

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Associate Professor Farzad Kamalabadi, Chair
Professor Erhan Kudeki
Professor Steven J. Franke
Associate Professor Minh N. Do

ABSTRACT

Accurate and efficient estimation of the key ionospheric state parameters such as electron density, ion composition, and electron and ion temperatures is required to understand fundamental issues of terrestrial plasma physics such as redistribution of energy and momentum, and coupling within terrestrial upper atmosphere regions. This work focuses on developing a modern coding scheme and inversion methodology using Arecibo incoherent scatter radar (ISR) to achieve efficient, and near-optimal estimates of the key ionospheric parameters. In particular, this work considers two aspects of the ISR inversion problem: (i) ISR lag estimates at individual altitudes, and (ii) modulation techniques that can provide more accurate estimates with a specific range resolution. These two aspects suggest a unifying framework for ISR inversion in which modern computational technology and ISR methodology are utilized in a robust estimation procedure.

The first contribution of this work is the development of a discrete forward model for F-region incoherent scatter measurements, where long pulses are utilized in transmission. The range smearing imposed on measurements by long pulses is modeled as a one-dimensional convolution along the range in the simplified case where the receiver sampling is instantaneous. The next major phase of this research is to develop an efficient hybrid technique that allows for estimation of plasma parameters by removing range smearing from measurements. The inversion technique incorporates both quadratic and edge-preserving regularization approaches in order to provide smooth plasma auto-correlation function (ACF) lag profiles in the presence of noise while still resolving the sharp gradients.

Another contribution of this thesis is to develop a technique for optimal modulation design in ISR experiments. The optimal resolution supported by ISR measurements is used as one criterion for the optimal design. The model order selection framework is applied to the problem at hand to find the optimal resolution. The results indicate that, compared with a long pulse, amplitude-

modulated codes yield finer range resolutions with nearly similar parameter estimation errors or smaller estimation errors with the same range resolution. In medium to high SNR scenarios, a smaller on-off ratio of the transmitted waveform is recognized as a determining factor for allowing more freedom in removing range ambiguities as well as resulting in improved statistical accuracy for integrated data in range and lag directions. In order to find the optimal amplitude modulation for Arecibo ISR measurements in medium to high SNR scenarios, a modified form of the sequential backward selection algorithm is applied to the space of all amplitude modulated pulses with a certain on-off ratio. Due to the vast search space, there is no possibility for an exhaustive search. Therefore, the problem of finding the optimal amplitude modulation is viewed as an optimization problem. Three optimality criteria, namely, sum of squared errors, uniformity of estimation errors, and condition number of convolution matrices, are considered.

The final contribution of this work was implementing and conducting several experiments in April 2004, August 2005, and July 2006 using the incoherent scatter radar at Arecibo Observatory, and applying the inversion technique to estimate the plasma parameters. In these experiments, the original mode of MRACF was modified to utilize amplitude modulation. The results of all these experiments verify that when the SNR is sufficiently high, compared with an unmodulated long pulse, improved range resolution with nearly the same statistical accuracy is obtained when an amplitude modulation is utilized.

The results of the developed methodology and experimental design of this work can be extended to other incoherent scatter radars (such as Jicamarca radar in Peru, and advanced modular incoherent scatter radar in Alaska) to improve the estimation task in other altitude and latitude regions, and to extract many further ionospheric parameters such as electric field strength, conductivity, current, and neutral wind speed.

*To my father, my mother, and my sisters,
and
my husband, Hossein,
for their love and support*

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my adviser, Dr. Farzad Kamalabadi, for his guidance, understanding, and patience during my graduate studies at the University of Illinois. He encouraged me to grow not only as a scientist but also as an instructor and an independent thinker. I am also thankful to him for encouraging the use of correct grammar and consistent notation in my writings and commenting on countless revisions of this manuscript. My special thanks goes to Dr. Erhan Kudeki for getting my graduate research started on the right foot, and always being there to listen and give advice. I am deeply grateful to him for the long discussions that helped me sort out the technical details of my work. I am also thankful to Dr. Steven Franke and Dr. Minh Do for commenting on my ideas and helping me understand and enrich them.

I am deeply indebted to Dr. Michael Sulzer, Dr. Nestor Aponte, and Dr. Sixto Gonzalez for their knowledge and advice. Their valuable comments and directions not only helped me in my research, but also taught me precious lessons for my future academic career. I also thank the staff at the Arecibo Observatory for their support during my data collection.

Most importantly, none of this would have been possible without the incredible support of my parents, sisters, and husband. My family to whom this dissertation is dedicated, has been a constant source of love, concern, support, and strength all these years. My special thanks goes to my parents who made it possible for me to pursue and complete my PhD degree. I am also thankful to all my friends at the University of Illinois for their friendship and support.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1 Motivation	1
1.2 Context and research contribution	3
1.3 Dissertation overview	6
CHAPTER 2 PRELIMINARIES	7
2.1 The basics of incoherent scattering of radio waves	7
2.1.1 Variation of the spectrum as a function of parameters	8
2.2 Principles of pulsed radar operations, the soft-target radar equation, and the ambiguity function	9
2.3 Conventional coding schemes in incoherent scatter experiments	13
2.3.1 Long-pulse technique	13
2.3.2 Multi-pulse technique	13
2.3.3 Alternating codes	15
2.4 Current methodologies in incoherent scatter inversion	16
2.4.1 Statistical framework for incoherent scatter inversion	17
2.4.2 Height-by-height analysis	20
2.4.3 Full-profile analysis	21
CHAPTER 3 PROPOSED TECHNIQUE FOR INVERSION OF INCOHERENT SCATTER MEASUREMENTS	24
3.1 Forward model	25
3.2 Inverse model	27
3.2.1 Deconvolution	27
3.2.2 Nonlinear least-squares (NLLS) optimization	34
3.3 Extension to 2-dimensions	35
3.4 Numerical results	37
3.4.1 Hybrid technique versus conventional processing techniques	37
3.4.2 Quadratic versus edge-preserving regularization techniques	44
CHAPTER 4 INVESTIGATION OF CODING SCHEMES FOR ISR MEASUREMENTS	47
4.1 Error covariance estimates	48
4.2 Fundamental resolution in ISR measurements	52
4.2.1 Model order selection framework	52
4.2.2 Model order selection method	54
4.3 Numerical estimates of range resolution	57

CHAPTER 5	EXPERIMENTAL INVESTIGATION OF CODING SCHEMES	
	WITH ARECIBO ISR	67
5.1	F-region measurements at Arecibo	67
5.1.1	Multiple radar auto-correlation function	68
5.2	Modified MRACF experiments at Arecibo	69
5.3	Observations	70
5.3.1	July 2006 experiment	70
5.3.2	August 2005 experiment	75
5.3.3	April 2004 experiment	87
CHAPTER 6	OPTIMAL MODULATION IN F-REGION INCOHERENT	
	SCATTER EXPERIMENT DESIGN	99
6.1	Experiment design in view of subset selection	100
6.2	Statistical formulation	101
6.3	Optimality criteria	102
6.4	Sequential backward selection algorithm	103
6.5	Numerical results	104
CHAPTER 7	CONCLUSIONS	110
7.1	Modeling	110
7.2	Inversion technique	111
7.3	Experiment design	111
7.4	ISR experiments and parameter estimation results	112
7.5	Future directions	113
7.5.1	Arecibo topside mode	113
7.5.2	Advanced modular incoherent scatter radar	114
7.5.3	Jicamarca radar	115
REFERENCES		116

CHAPTER 1

INTRODUCTION

1.1 Motivation

Incoherent scatter radar (ISR) is the most powerful ground-based technique for studying the Earth's ionosphere. The incoherent scatter (IS) echo is the result of the scattering of electromagnetic energy, radiated from radar, by electrons in the ionospheric plasma, which are themselves controlled by much slower, massive positive ions. The frequency spectrum of the received signal provides information about electron and ion temperatures, electron density, ion composition and velocity. The analytical relationship between the spectrum and these parameters has been well established in the literature, e.g. [*Dougherty and Farely*, 1960, 1961; *Farley*, 1966, 1969; *Hagfors*, 1961, 1971]. With the extraction of these parameters from incoherent scatter measurements, one can deduce many further ionospheric parameters such as ion composition, electric field strength, conductivity and current, neutral air temperature and wind speed.

Although the exact forward theory of incoherent scatter was established more than four decades ago, inversion, the estimation of parameters from incoherent scatter spectra, has remained an open-ended problem, due to two major factors. The first complication stems from the fact that variation of different plasma parameters may give rise to similar changes in the IS spectrum [*Vallinkoski*, 1988], e.g. the distinction between changes of the spectra due to ion composition or temperature ratio is very difficult. The same is true for ion mass and ion temperature for fixed temperature ratio. This implies that spectrum, or equivalently its Fourier transform, i.e. auto-correlation function (ACF), needs to be retrieved as accurately as possible, since small errors in ACF may yield large errors in estimated parameters. The second factor is the range smearing of information from one altitude over a number of altitudes, which is due to the length of the transmitted pulse.

The transmitted pulse should be long enough to provide sufficient spectral resolution in measuring the spectrum, or equivalently, to permit ionospheric ACF to be measured with sufficient lag extent. On the other hand, at any given time, the received signal includes contributions from a volume which extends $\frac{cT}{2}$ in the radar line-of-sight direction, where c is the speed of light and T is the length of the pulse.

An example of the effect of range smearing is shown in Figure 1.1 where different lags of the ACF are plotted prior to and after imposing the range ambiguity for an un-coded modulation of length $280 \mu\text{s}$ with a 430 MHz radar (thin and thick curves, respectively). The estimation of ionospheric parameters from the range-smeared ACF would result in a greater ion temperature than electron temperature, which is not physical [Nikoukar *et al.*, 2008]. Conventional processing techniques utilize various ways to tackle the inclusion of range smearing into analysis. Range-gate analysis, for example, is based on the assumption of constant plasma parameters at each range gate (distance covered by the pulse length) and compensates for the effect of range smearing by a simple triangular weighting. This method, although simple and fast, is based on an unrealistic assumption and suffers from the coarse resolution of estimated parameters. Full-profile analysis, which attempts to estimate ionospheric parameters at all altitudes simultaneously, on the other hand, considers a full model for range smearing. This technique, however, suffers from its significant computational cost, which can be mitigated by coarser interpolation of the parameter space, hence sacrificing estimation accuracy. The limitations of these processing techniques have motivated the search for more robust approaches to the problem of ISR inversion that meet the challenges of modern ISRs.

The goal of this research is to develop a modern ISR parameter estimation methodology to achieve efficient, near-optimal estimates of ionospheric parameters with a fine resolution. Primarily, this research will focus on the F-region ionosphere over the Arecibo Observatory, where the effect of molecular ions and collision can be simply neglected in the analysis. Moreover, the beam direction over Arecibo can be far from perpendicular to the magnetic field lines, where the incoherent scatter process is well known. This work will consider two aspects of the ISR inversion problem: (i) ISR lag estimate at individual altitudes and (ii) modulation techniques that can provide more accurate estimates with a specific range resolution. These two aspects suggest a unifying framework for ISR inversion in which modern computational technology and ISR methodology are utilized in a robust estimation procedure.

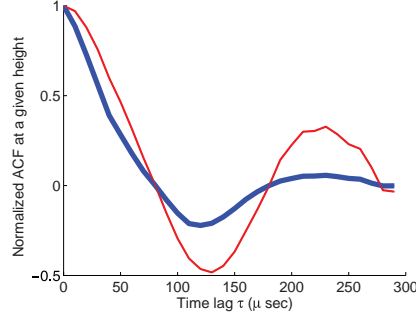


Figure 1.1 Normalized theoretical and measured ACF for a long un-coded pulse transmission and a constant ionosphere (thin and thick curves, respectively). Range smearing makes the shape of ACF so distorted that results of the fit by standard nonlinear least-squares methods become poor. Notice that the above ACFs have been generated using a radar frequency of 430 MHz for a single ion ionosphere (oxygen). The lag spacing is $10 \mu\text{s}$ and an un-coded long pulse of length $280 \mu\text{s}$ is considered for transmission.

1.2 Context and research contribution

Since the development of ISR theory and construction of large incoherent scatter radars, such as the Arecibo radar, various methodologies were invented to deal with difficulties attributed to ISR parameter estimation. Generally speaking, they fall into two major categories of analysis techniques and coding schemes. To date numerous coding schemes have been primarily designed to make direct measurements of plasma ACF possible despite the pulse length. There are two main classes of coding techniques, amplitude modulation and phase modulation schemes, used for transmission. Amplitude modulation schemes such as double-pulse and multi-pulse techniques, pioneered in [Farley, 1969, 1972], exploit short pulses to obtain the spectra at individual altitudes. They, however, suffer from great sensitivity to the background and receiver noise. These techniques are suitable only for parameter estimation around the F-region peak where the back-scattered signal is rather high. Phase modulation schemes, such as alternating codes, maintain the fine spatial resolution of measurements, while exploiting the full duty cycle of radar, and are not as sensitive to the noise level. They, however, are only appropriate for low signal-to-noise ratio (SNR) scenarios,¹ where the lag estimate errors are independent from each other. As a result, in the case of the Arecibo radar, where the transmitted power is high and the transmitting and receiving antenna area is large, this technique loses its appeal over other existing coding techniques.

¹The power of zeroth lag divided by the background noise power in a short pulse transmission.

Currently, the usual mode of operation for this type of measurements at Arecibo Observatory is the simple long-pulse technique. The major difficulty associated with the long-pulse modulation is that it can only provide range-smeared (indirect) measurements of plasma ACF. The current analysis techniques developed to deal with this range smearing of information suffer from severe sensitivity to noise, coarse spatial resolution, or significant computational expense. Although the relationship between the measured signal ACF and the plasma ACF at individual altitudes has been thoroughly derived in the work of *Lehtinen and Huuskonen* [1996], to date no systematic approach exists to exploit this relationship in favor of more efficient methodologies. In this work, we present an efficient, near-optimal approach to the problem of incoherent scatter radar inversion which is extremely robust in the presence of system and background noise. The key to this new method is to consider the system of measurements as a combination of two linear and nonlinear systems. We formulate the estimation of plasma parameters as the result of linear inversion of measured signal ACF (to achieve plasma ACF at individual altitudes) and then of nonlinear inversion of plasma ACF (to obtain plasma parameters). Although the proposed analysis approach attempts to obtain estimates of plasma ACF at individual altitudes (separated by the distance equal to lag resolution), it is only able to partially remove the smearing; and thus there is a limit to the resolution of the estimated physical parameters.

Another drawback associated with un-coded long-pulse measurements is that they suffer from high correlation between ACF lag estimate errors. This correlation plays an important role in ISR error analysis, that is, determining the uncertainty level in estimated parameters. First, in order to avoid underestimation of error bars, one has to account for the full error covariance matrix.² In addition, when SNR is high, random fluctuations in the ACF estimates originate in part from the self-noise, which is generated by the uncorrelated scatter from the transmitted signal from all simultaneously illuminated altitudes [*Lehtinen and Huuskonen*, 1996]. When the self-noise is dominating, the accuracy of observations cannot be increased by further increasing the signal strength because the random error in data increases at the same rate. Another effect of the correlation between lag estimate errors is related to the range integration. Because the estimates are highly correlated in high SNR regimes, the addition of N ACFs from adjacent altitudes does

²An error covariance matrix is formed from the correlation between errors of different lag estimates received at different time indices.

not yield an \sqrt{N} -fold increase in estimation accuracy as would be the case for fully independent observations.

In this work we present amplitude modulation (AM) as the solution to reducing the correlation between ISR lag estimate errors. This reduction in correlation, in turn, allows for improving range resolution especially in high SNR regimes. For this purpose, we study and compare the performance of amplitude modulation with that of long-pulse modulation in terms of statistical accuracy and range resolution by means of both numerical simulations and real data analysis. We show that the correlation between lag estimate errors is highly affected by the envelop of the modulated waveform for a given SNR, and that amplitude modulation reduces this correlation significantly. We show that this lower correlation results in improved accuracy when measurements are integrated in range and lag. Using model order selection methods, we also derive a mathematical measure for estimating the fundamental range resolution supported by ISR measurements.

Moreover, we present parameter estimation results of several F-region experiments that we conducted with Arecibo radar in 2004, 2005, and 2006. In these experiments, long-pulse and various amplitude modulation codes were alternated in transmission to make a direct comparison between the estimation results possible. The on-off ratio of the transmitted waveform is represented as a major determining factor in the performance of an amplitude modulated code. However, slight differences in performances of AM pulses with the same on-off ratio necessitate consideration of other related factors, such as uniformity of the number of integrated heights for different lags, or condition number of convolution matrices. In this work, we present close-to-optimal waveform design for ISR experiments using subset selection methods. Our method, however, provides preliminary results and its convergence properties have yet to be developed.

The developed ISR methodology can also be extended to other incoherent scatter radars such as Jicamarca radar in Peru or advanced modular incoherent scatter radar (AMISR) in Poker Flat, Alaska. Although the requirements of the plasma correlation measurements for these radars are quite different from those for the Arecibo radar, the developed methodology can be expanded to meet the requirements.

1.3 Dissertation overview

This dissertation is organized, as follows: Chapter 2 presents an overview of the incoherent scatter theory, the incoherent scatter radar equation, and the concepts of range smearing and ambiguity. It also covers the conventional coding schemes and conventional incoherent scatter data processing techniques. The first contribution of this work is the development of a hybrid technique that allows for efficient estimation of plasma parameters from ISR measurements. In Chapter 3, a technique is developed which combines regularization theory and deconvolution methods. Two different regularization methods, quadratic and edge-preserving methods, are considered. A statistical interpretation of the proposed algorithm is given using statistical formulation. Numerical results are presented.

Another contribution, presented in Chapters 4, 5, and 6, is the development of a framework for determining the fundamental resolution, and optimal experiment design, in ISR measurements. We also compare the performances of long-pulse and amplitude modulation, which results in lower correlation in lag estimate errors. Numerical estimates of ACF lag variance reduction are presented for a long pulse and several amplitude modulated pulses when measurements are integrated in range and lag directions. Moreover, the improvement in range resolution for estimated parameters is presented for AM pulses using numerical simulation in Chapter 4. Chapter 5 describes several ISR experiments conducted using the radar at Arecibo observatory, and presents a compelling comparison between the performances of long pulse and AM pulses with different on-off ratios, as well as AM pulses with the same on-off ratio. In Chapter 6, optimal design of the transmitted waveform in F-region ISR experiments is considered, and a practical algorithm is proposed for finding the optimal pulse configuration.

Finally, Chapter 7 provides concluding remarks, summarizes the dissertation's contributions, and discusses some future research directions.

CHAPTER 2

PRELIMINARIES

2.1 The basics of incoherent scattering of radio waves

The idea of incoherent scatter radars was initiated by W. Gordon in 1958. He suggested that if a powerful beam of radio waves with a frequency well above the plasma frequency was sent vertically through the ionosphere, an extremely small but still measurable amount of power would scatter back to the ground by small-scale random fluctuation of electrons in the ionosphere. The backscattered signal could be roughly estimated as the sum of the powers scattered from each electron, as given by the Thompson electron cross-section [Kudeki, 2003; Nygren, 1996]. Since the electrons are randomly distributed in the ionosphere, they scatter incoherently, hence the name “incoherent scatter,” and the power scattered by each is summed to obtain the total signal. Moreover, owing to the thermal motions of the electrons, the scattered radiation would be subject to a Doppler broadening, with a frequency deviation of the order of $(KT/m_e)^{1/2}/\lambda_0$, where K is the Boltzmann constant, T is the temperature, m_e is the electron mass, and λ_0 is the radar wavelength.

The first successful ISR experiment was performed by Bowles in 1958 using extremely high-powered equipment. His results indicated that the returned power was of the order of the power predicted by Gordon; however, the observed Doppler broadening was much smaller than what Gordon had anticipated. Different groups of scientists investigated this discrepancy [Dougherty and Farely, 1960, 1961; Farley, 1966; Hagfors, 1961, 1971]; and although they used different theoretical approaches, they all arrived at the following conclusion: Provided that the Debye length (the distance over which the influence of the electric field of an individual charged particle is felt by the other charged particles inside the plasma [Bittencourt, 2004]) is sufficiently large, the ions play an important role in determining the Doppler spread, although it is the electrons that do the actual

scattering and that the effect of ions in the form of Coulomb collisions must be considered in the electron scattering.

In the next section, we describe the basic idea behind the incoherent scatter of the radio waves with pulsed incoherent scatter radars, and elaborate on the expression of the incoherent scatter spectrum, or equivalently its Fourier transform, auto-correlation function, as a function of ionospheric parameters. We then demonstrate the effect of varying parameters on the resulting spectra or ACFs by a few examples.

2.1.1 Variation of the spectrum as a function of parameters

In the presence of a magnetic field (\mathbf{B}) in a plasma with Coulomb interactions and a beam direction away from perpendicular to \mathbf{B} , the spectrum of the electron density fluctuation, $\langle |n_e(\mathbf{k}, \omega)|^2 \rangle$, can be represented by

$$\langle |n_e(\mathbf{k}, \omega)|^2 \rangle = 2N_0 \frac{|j(k^2 h_e^2 + \mu) + \mu \theta_i J(\theta_i)|^2 \frac{Re\{J(\theta_e)\}}{k\sqrt{2}C_e} + |j + \theta_e J(\theta_e)|^2 \frac{Re\{J(\theta_i)\}}{k\sqrt{2}C_i}}{|j(k^2 h_e^2 + 1 + \mu) + \theta_e J(\theta_e) + \mu \theta_i J(\theta_i)|^2} \quad (2.1)$$

where $\langle . \rangle$ denotes the expectation operation, N_0 represents electron density, μ is the electron-ion temperature ratio T_e/T_i , h_e is the Debye length, and θ_s represents the normalized velocity of the particle s . The Gordyev-type integral $J(\theta_s)$ is described by

$$J(\theta_s) = \int_0^\infty dt e^{-j\theta t} e^{-[\frac{t^2}{4} \sin^2 \alpha + \frac{1}{\phi_s^2} \sin^2(\frac{\phi_s t}{2}) \cos^2 \alpha]} \quad (2.2)$$

where α is the magnetic aspect angle (the complement of the angle between propagation vector \mathbf{k} and magnetic field \mathbf{B}), and ϕ denotes the normalized gyrofrequency of the species s (Ω_s), i.e. $\phi_s = \frac{\Omega_s}{k\sqrt{2}C_s}$ (see [Kudeki, 2003] for a complete discussion).

In thermal equilibrium where $T_e = T_i$, the spectrum consists of two peaks close to the origin with widths corresponding to ion thermal velocities. Therefore, the spectrum of a plasma has a small dip at the central frequency corresponding to no Doppler shift. In order to provide a qualitative picture of the effect of various parameters on the spectrum (and its Fourier transform equivalent, ACF), spectra (and ACFs) at various parameter combinations are plotted in Figure 2.1 (a) and (b). Plots

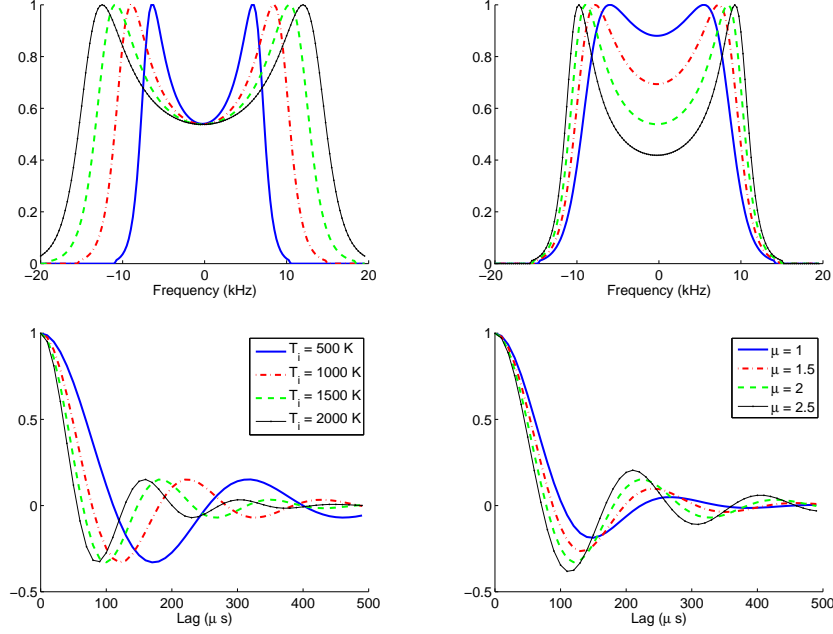


Figure 2.1 Variation of the ISR spectrum and ACF as a function of ion temperature (left panels) and temperature ratio (right panels), respectively.

in the left panels show the variation of the spectrum and ACF as a function for fixed temperature ratio ($\mu = 2$) and as a function of ion temperature, whereas plots in the right panels show the variation of the spectrum and ACF for fixed ion temperature ($T_i = 1000$ K) and as a function of temperature ratio. All plots have been generated for frequency of 430 MHz and a 100% oxygen ionosphere. As seen from the plots, for a fixed μ , the ion line becomes broadened as ion temperature increases. Moreover, the peak-to-valley ratio also increases with increasing temperature ratio with ion temperature fixed.

In the next section, we describe how we can infer the plasma spectrum or ACF from an incoherent scatter radar.

2.2 Principles of pulsed radar operations, the soft-target radar equation, and the ambiguity function

In order to introduce the soft-target radar equation, we first provide a quick overview of the principles of pulsed radar ISR operation. For this purpose, let us denote the envelope and the length

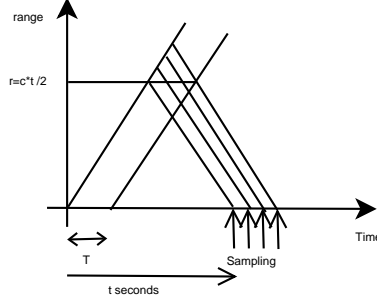


Figure 2.2 Transmission and reception scenario in a typical incoherent scatter radar. A pulse (modulated or unmodulated) of length T is transmitted through the ionosphere, and the return is sampled at δt intervals.

of the transmitted waveform in the radar experiment by $V_0(t)$ and T , respectively. Suppose the pulse is a baseband waveform modulated by a sinusoidal wave of frequency ω_0 . The corresponding wavelength and wave vector will be $\lambda_0 = \frac{2\pi c}{\omega_0}$ and $\mathbf{k}_0 = \frac{2\pi}{\lambda_0} \hat{\mathbf{k}}$, respectively, where c is the speed of light and $\hat{\mathbf{k}}$ is the direction of wave transmission. The returned signal is then sampled with a period δt (Figure 2.2) [Nikoukar *et al.*, 2008].

The scattered signal from the electron density fluctuations in the ionosphere can be represented by [Kudeki, 2003]

$$v(t) \propto \int_{\mathbf{r}} n(\mathbf{r}, t) e^{-j2k_0 r} V_0(t - \frac{2r}{c}) d\mathbf{r} \quad (2.3)$$

where $n(\mathbf{r}, t)$ represents the fluctuating component of the plasma electron density at range $\mathbf{r} = r\hat{\mathbf{r}}$ along the direction $\hat{\mathbf{r}}$. The proportionality factor, $Z(\mathbf{r})$, describes all geometrical effects and can be written as:

$$Z(\mathbf{r}) = R_{rad} \frac{G(\hat{\mathbf{r}}) r_e}{2k_0 r^2} \quad (2.4)$$

where $r_e \approx 2.8 \times 10^{-15} \text{m}$ is the classic electron radius, R_{rad} represents the radiation resistance of the antenna, and $G(\hat{\mathbf{r}})$ is the antenna gain along the direction $\hat{\mathbf{r}}$. Note that the received signal spectrum is centered at the carrier frequency, ω_0 , instead of zero, as the antenna output is modulated by the carrier signal. At the coherent detector, the input is mixed with the signal derived from a local oscillator centered at the carrier frequency and is low-pass filtered. This task brings the output signal to the baseband again, centered at zero frequency. The exponential term in (2.3) represents

this mixing operation.

Since there are many scattering electrons in the ionosphere, the scattered signals (or received voltages) can be treated as Gaussian random variables to a very good approximation, according to the central limit theorem [Papoulis, 1986]. Therefore, instead of working directly with voltage samples, we model their joint statistics in the form of the auto-correlation function, $\langle v(t)\overline{v(t+\tau)} \rangle$, where τ represents the time lag and $\overline{(\cdot)}$ denotes the conjugation operation, respectively.

From (2.3), the ACF of the received voltage can be expressed as

$$\langle v(t)\overline{v(t+\tau)} \rangle \propto \int_{\mathbf{r}'} \int_{\mathbf{r}} \langle n(\mathbf{r}, t) \overline{n(\mathbf{r}', t+\tau)} \rangle V_0(t - \frac{2r}{c}) \overline{V_0(t + \tau - \frac{2r'}{c})} e^{-2jk_0(r'-r)} d\mathbf{r} d\mathbf{r}' \quad (2.5)$$

where $\mathbf{r}' = r'\hat{\mathbf{r}}'$ denotes the range r' along the direction $\hat{\mathbf{r}}'$. Within the integrand above $\langle n(\mathbf{r}, t) \overline{n(\mathbf{r}', t+\tau)} \rangle$ is the space-time ACF of the density fluctuation $n(\mathbf{r}, t)$. With the assumption that fluctuations have homogenous and stationary statistics which vanish rapidly with an increasing magnitude of $\mathbf{x} \equiv \mathbf{r}' - \mathbf{r}$, we can proceed as

$$\langle v(t)\overline{v(t+\tau)} \rangle \propto \int_{\mathbf{r}} V_0(t - \frac{2r}{c}) \overline{V_0(t + \tau - \frac{2r}{c})} d\mathbf{r} \int_{\mathbf{x}} \langle n(\mathbf{r}, t) \overline{n(\mathbf{x} + \mathbf{r}, t+\tau)} \rangle e^{-2j\mathbf{k}_0 \cdot \mathbf{x}} d\mathbf{x} \quad (2.6)$$

The inner integral represents a spatial Fourier transform, denoting the ACF of electron density fluctuations at altitude r , time lag τ , and the wave vector ($\mathbf{k} = -2\mathbf{k}_0$), as follows:

$$R(\mathbf{k}, r, \tau) \equiv \int_{\mathbf{x}} \langle n(\mathbf{r}, t) \overline{n(\mathbf{x} + \mathbf{r}, t+\tau)} \rangle e^{-2j\mathbf{k}_0 \cdot \mathbf{x}} d\mathbf{x} \quad (2.7)$$

Notice that this ACF of electron density fluctuations is the same as the plasma ACF whose analytical expression (as a function of radar and plasma parameters) was already introduced in the previous section.

By inserting (2.7) into (2.6) we obtain [Kudeki, 2003]:

$$\langle v(t)\overline{v(t+\tau)} \rangle \propto \int_{\mathbf{r}} V_0(t - \frac{2r}{c}) \overline{V_0(t + \tau - \frac{2r}{c})} R(\mathbf{k}, \mathbf{r}, \tau) d\mathbf{r} \quad (2.8)$$

The above equation which demonstrates the relationship between the received voltage and the target in a statistical sense is called the soft-target radar equation and provides the necessary background for estimation of ionospheric parameters. Inspection of this equation brings us to the following two observations:

1. The incoherent scatter radar equation holds for each lag of ionospheric ACF independently from other time lags. Hence, to evaluate the plasma ACF at time lag τ , we need to consider only the same lag of the received signal ACF.
2. Using incoherent scatter radar, it is not possible to achieve point estimations of the plasma ACF. Instead, weighted averages of this quantity over a finite-range interval are obtained. The weights are merely dependent on the modulated waveform and vary from lag to lag.

The functions describing the averaging operation on the underlying plasma ACF are called soft-target radar ambiguity functions, $(\mathbf{p}_\tau(t) = V_0(t)\overline{V_0(t+\tau)})$. These functions essentially indicate that the signal coming from range r contains information from several altitudes, where the altitude interval is equal to the distance covered by the product of the waveform and its shifted version. Note that the zeroth lag contains the most range smearing of information. As we move to farther lags, this altitude interval decays as the common part between the pulse and its shifted version diminishes.

Notice that in a more general case, the ambiguity functions are dependent on both range variables and time lag, as developed in *Lehtinen and Huuskonen* [1996] and *Holt et al.* [1992]. This dependence on time lag is caused by a non-ideal receiver, whose impulse response contains the time average of previous samples. For now, however, we assume the receiver has a sufficiently narrow impulse response, and as such the ambiguity in the lag direction is negligible. The expansion of the radar ambiguity function to the 2-dimensional case is considered in Section 3.3.

Also note that our definition of the soft-target ambiguity function differs from that of hard-target radar applications where it is defined as (see for example [Blahut, 2004])

$$\chi(\tau, \nu) = \int V_0(t)\overline{V_0(t+\tau)}e^{-j2\pi\nu t}dt \quad (2.9)$$

In this equation, $\chi(\tau, \nu)$ is the hard-target ambiguity function in terms of a delay (range) resolution

variable, τ , and doppler resolution, ν . The hard-target radar equation is used as a measure of detectability of two hard targets with separation in range and velocity. The difference between the two definitions is due to different filters being used in radar receivers. In a hard-target application, normally a matched filter is used, whereas in soft-target radars, a receiver with a boxcar impulse response of short length is often utilized. Through the rest of this work, we use the term “ambiguity function” to refer to soft-target radar applications only.

2.3 Conventional coding schemes in incoherent scatter experiments

In this section, we describe the conventional coding schemes that are used in typical F-region incoherent scatter experiments. These modulation techniques include long-pulse and multi-pulse techniques, and alternating codes.

2.3.1 Long-pulse technique

In long-pulse transmission, let $V_0(t) = 1$ for $0 \leq t < T$, where T represents the length of the pulse. In this case, the one-dimensional ambiguity function $p_\tau(t)$ can be represented as follows:

$$p_\tau(t) = \begin{cases} 1 & \text{if } 0 \leq t < T - \tau \\ 0 & \text{otherwise} \end{cases}$$

In other words, the ambiguity function is a square pulse whose width decreases as the time lag increases. Therefore, in long-pulse transmission, it is not possible to retrieve point estimations of the plasma ACF directly from the data. To do so, one needs to account for the range ambiguity and has to devise methods that are described in the next section as well as the next chapter.

2.3.2 Multi-pulse technique

The class of multi-pulse codes pioneered by *Farley* [1972] is a set of amplitude modulated pulses, which can achieve point estimation of ionospheric ACF (nonzero) lag estimates at individual alti-

tudes without ambiguity. A classical pulse-code modulation meets the following conditions:

1. The code consists of N pulses of equal duration (bit length).
2. The distance of any pair of pulses is an integer multiple of the smallest (basic) inter-pulse distance (τ), and all distances are different. The lag resolution is set to be equal to the basic inter-pulse distance.

Although the multi-pulse technique can provide the ionospheric ACF estimation at individual altitudes, its major drawback is its low duty cycle, and thus its high sensitivity to the background noise level. For example, a six-pulse code with a bit length of $17 \mu\text{s}$, will have a total length of $306 \mu\text{s}$ (see Figure 2.3). This configuration yields a duty cycle of only 33%. As the number of short pulses increases, the duty cycle decreases. For example, a 7-pulse code has a duty cycle of 26% only. In order to facilitate the full radar duty cycle, and therefore to reduce the sensitivity of the method, the technique is usually implemented in the form of interlaced codes. In an experiment with these interlaced codes, the transmitted frequency is changed rapidly so that gaps are not left in the transmission. Another drawback of the multi-pulse codes is that the zero lag measured by this technique is normally discarded. The reason is that the shape of the ambiguity function for the zeroth lag differs from that of others, and its corresponding scattering volume consists of a number of individual volumes instead of a single volume. The use of multi-pulse zero lag data, however, can be used to improve incoherent scatter radar power profile accuracy, as suggested by *Lehtinen and Huuskonen* [1986].

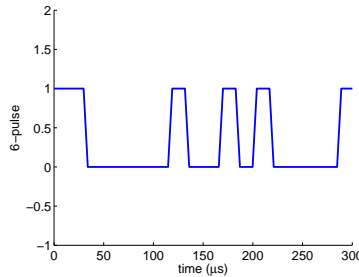


Figure 2.3 Six-pulse technique. The baud is $17 \mu\text{s}$ and the pulses are located at 0, 17, 136, 187, 221, and $289 \mu\text{s}$.

2.3.3 Alternating codes

Alternating codes are a series of phase-coded pulses which are transmitted one at a time through the ionosphere, and once all the pulses are transmitted, the cycle starts over again. Each single pulse consists of a combination of elementary pulses (bauds) with signs ± 1 . These signs are changed from pulse to pulse in a way that the ambiguity function due to all of these pulses has a single peak on a zero background.

The working principle of this coding method is shown in Figure 2.4. The matrix on the left shows a set of four phase-coded pulses (scan count), where each pulse consists of four elementary pulses with varying signs. The matrix marked by W_1 denotes the sign of the ambiguity functions of the first lag. This is obtained by multiplying each column of the matrix by its adjacent one. The matrices in the square region are obtained by further multiplying each column of W_1 by the first through third column. According to the figure, all but one of the columns of each matrix sum to zero. Recalling that each column in the original envelope of the pulses corresponds to one particular altitude, one can easily see that the undesired signals from the other two altitudes have been canceled out, while the signal from the altitude of interest is preserved. The same argument holds for other lags of the ambiguity function, W_2 and W_3 . The only difference is that as we consider higher lags, the number of columns (altitudes) decreases, until a single column is left for W_3 .

In their work, *Lehtinen and Huuskonen* [1996] have presented an analytical solution to the problem of choosing signs of every bit in all scans such that the whole set shows the desirable properties of single peaks for the ambiguity functions. For this purpose, they exploit one of the properties of the Hadamard matrices, which is orthogonality of rows [Harwit and Sloane, 1979]. As a result, certain arrangements of the rows or columns achieve the analytical cancelations. They have also shown that in situations where the received impulse response has a finite width, as opposed to a δ -like response, the required number of scans in each cycle is twice the number of elementary pulses in each scan. This requirement reduces the time resolution of the measurements by a factor of $1/N_p$, where N_p is the number of scan counts in one complete cycle.

Although alternating codes guarantee the exact cancelation of the signal contribution from unwanted regions, they fail to remove the covariances between the lag-estimate errors in an ISR

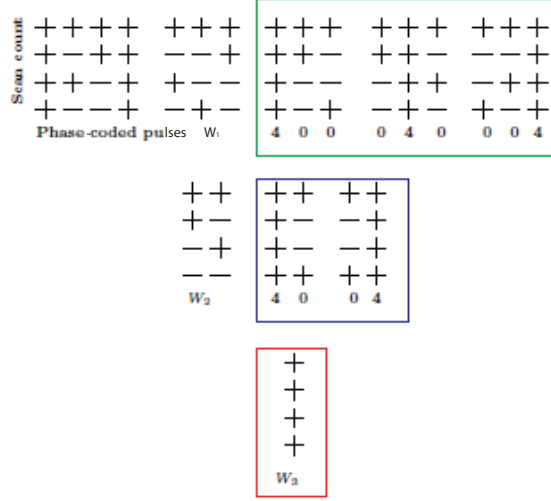


Figure 2.4 Working principle of the alternating codes. Ambiguity functions are shown for all possible three lags (W_i , where i is the lag number). The matrices next to each ambiguity matrix are formed by multiplying each column (altitude) of this matrix by other columns (altitudes). This multiplication results in the canceled contributions from these columns, while preserving the signal from the first altitude.

experiment. Covariance calculations, however, require the lag profile matrices from all the codes, not only the final results. This requirement makes the method even more computationally expensive, but makes it more efficient when the SNR is very low and the lag estimate errors are nearly independent.

2.4 Current methodologies in incoherent scatter inversion

Let us assume that one of the above-mentioned coding schemes has been utilized in an ISR experiment, and estimates of the received signal ACF lags (or plasma ACF lags provided that multi-pulse or alternating codes are used) are available. In this section we provide an overview of the statistical framework for inversion of incoherent scatter radar measurements to obtain estimates of ionospheric parameters. We formulate the problem of inversion in terms of the maximum likelihood (ML) principle. Next, we explain conventional inversion methodologies, height-by-height analysis and the full-profile technique.

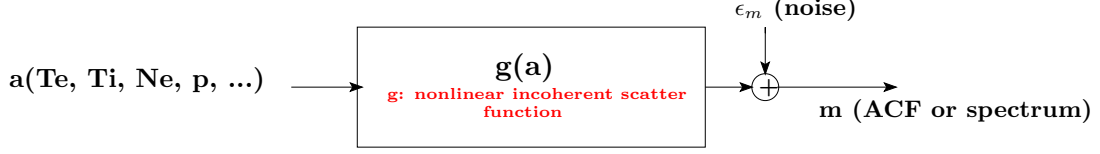


Figure 2.5 Incoherent scatter process as a hypothetical system, where a which is a vector of plasma parameters is the input of the system, m , the plasma spectrum or ACF, is the output, and the nonlinear function of incoherent scatter, which relates the parameters to plasma spectrum, is denoted by $g(\cdot)$.

2.4.1 Statistical framework for incoherent scatter inversion

Let us consider the incoherent scatter process as a system, shown in Figure 2.5 and denoted by $g(\cdot)$, in which the ionospheric state parameters form the input, \mathbf{a} , and the spectrum or ACF of the received signal forms the output, \mathbf{m} . The relationship between the measurements and the desired parameters can be written as

$$\mathbf{m} = g(\mathbf{a}) + \epsilon \quad (2.10)$$

where ϵ denotes the noise of the system and in the most general case is signal dependent. However, it gains Gaussian characteristics as the measurements are integrated over many transmissions. Therefore, it can be well described by its mean, $\mu = \langle \epsilon \rangle$, and covariance matrix $\Sigma_\epsilon = \langle \epsilon \epsilon^T \rangle$. A nonzero mean would indicate biases, and without loss of generality can be considered as zero.

The conditional probability density function now can be expressed by

$$p(\mathbf{m}|\mathbf{a}) = \frac{1}{(2\pi)^{n/2} |\Sigma_\epsilon|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{m} - g(\mathbf{a}))^T \Sigma_\epsilon^{-1} (\mathbf{m} - g(\mathbf{a})) \right) \quad (2.11)$$

where n is the number of data points. The joint probability density function of \mathbf{m} and \mathbf{a} can be described as

$$\begin{aligned} p(\mathbf{m}, \mathbf{a}) &= p(\mathbf{a}) p(\mathbf{m}|\mathbf{a}) \\ &\sim p(\mathbf{a}) \frac{1}{(2\pi)^{n/2} |\Sigma_\epsilon|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{m} - g(\mathbf{a}))^T \Sigma_\epsilon^{-1} (\mathbf{m} - g(\mathbf{a})) \right) \end{aligned} \quad (2.12)$$

Let us now suppose that the *a priori* density of the parameters ($p(\mathbf{a})$) is approximately constant in the regions where the conditional probability density function is significantly different from zero.

The ML estimate is then equivalent to minimizing the quadratic form in the exponent in (2.12), i.e. $\chi^2 = -\frac{1}{2}(\mathbf{m} - g(\mathbf{a})^T \Sigma_\epsilon^{-1}(\mathbf{m} - g(\mathbf{a})))$. Thus, under the assumption of Gaussian errors, the ML estimate can be obtained via the quadratic or least-squares optimization procedures. If the matrix Σ_ϵ is diagonal, this quadratic form is reduced to the sum of the differences between the components of the predicted data, $g(\mathbf{a})$, and the components of the measurements weighted by the inverses of the diagonal elements of Σ_ϵ . On the other hand, in situations where Σ_ϵ is not diagonal, the quadratic form of (2.12) defines a generalized least-squares estimation method.

Linear approximation

Let us suppose that the correct values of the parameters are \mathbf{a}_0 and the predictive measurements are denoted by $\mathbf{m}_0 = g(\mathbf{a}_0)$. If the errors are small, the solution to the least-squares problem is very close to \mathbf{a}_0 , and as such we can use the Taylor expansion of the first degree to state the nonlinear quadratic form of (2.12)

$$\mathbf{m} - \mathbf{m}_0 = g(\mathbf{a}) - g(\mathbf{a}_0) + \epsilon = \mathbf{A}(\mathbf{a} - \mathbf{a}_0) + \epsilon \quad (2.13)$$

where $\mathbf{A}_{ij} = \partial g_i / \partial a_j$ is the partial derivative of the incoherent scatter function with respect to plasma parameters. Thus the errors derived are only valid in the limit of small fluctuations in the measurements. In practice this condition can be satisfied by using longer integration times. Replacing $\mathbf{a} - \mathbf{a}_0$ and $\mathbf{m} - \mathbf{m}_0$ by \mathbf{a} and \mathbf{m} , respectively, yields

$$\mathbf{m} = \mathbf{A}\mathbf{a} + \epsilon \quad (2.14)$$

The linear formulation is necessary and useful because the likelihood or posteriori distribution, and thus the solution, can be expressed in analytical forms. Here, we focus on the ML estimate where, with the assumption of the Gaussian noise, the likelihood can be described as

$$p(\mathbf{m}|\mathbf{a}) \sim \exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{A}\mathbf{a})^T \Sigma_\epsilon^{-1}(\mathbf{m} - \mathbf{A}\mathbf{a})\right) \quad (2.15)$$

In order to bring the expression for the transitional density to a standard quadratic form, we first consider the Cholesky decomposition of the inverse noise covariance matrix as $\Sigma_\epsilon^{-1} = \mathbf{D}^T \mathbf{D}$ (since

Σ_ϵ is positive semi-definite). We proceed by simplification of (2.15) and using the following identity $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \mathbf{I}$

$$\begin{aligned}
p(\mathbf{m}|\mathbf{a}) &\sim \exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{A}\mathbf{a})^T \mathbf{D}^T \mathbf{D} (\mathbf{m} - \mathbf{A}\mathbf{a})\right) \\
&= \exp\left(-\frac{1}{2}(\mathbf{D}\mathbf{m} - \mathbf{D}\mathbf{A}\mathbf{a})^T (\mathbf{D}\mathbf{A}[(\mathbf{D}\mathbf{A})^T(\mathbf{D}\mathbf{A})]^{-1}(\mathbf{D}\mathbf{A})^T)(\mathbf{D}\mathbf{m} - \mathbf{D}\mathbf{A}\mathbf{a})\right) \\
&= \exp\left(-\frac{1}{2}(\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{m} - \mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A}\mathbf{a})^T (\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A})^{-1} (\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{m} - \mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A}\mathbf{a})\right)
\end{aligned} \tag{2.16}$$

Factoring the term $\mathbf{Q} = \mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A}$ and further simplification yields

$$p(\mathbf{m}|\mathbf{a}) \sim \exp\left(-\frac{1}{2}(\mathbf{Q}^{-1} \mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{m} - \mathbf{a})^T \mathbf{Q} (\mathbf{Q}^{-1} \mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{m} - \mathbf{a})\right) \tag{2.17}$$

The ML estimate, which minimizes the likelihood, can be analytically expressed as

$$\mathbf{a}_{ML} = \mathbf{Q}^{-1} \mathbf{A} \Sigma_\epsilon^{-1} \mathbf{m} \tag{2.18}$$

where its error covariance matrix is given by

$$\mathbf{Q}^{-1} = (\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A})^{-1} \tag{2.19}$$

The matrix \mathbf{Q} is called the Fisher information matrix [Poor, 1994]. Clearly, all the measurement components make certain contribution to the Fisher information matrix.

The statistical inversion as described above is applicable to ISR inversion only in situations where the plasma ACF is readily available, by utilizing multi-pulse or alternating codes, for example. In cases where the modulation does not provide ambiguity-free measurements, one has to account for the effect of ambiguity in some way. Below, we review two of such methods, namely, height-by-height analysis and the full-profile technique. In the next chapter, we introduce a hybrid technique for the inversion of ISR measurements.

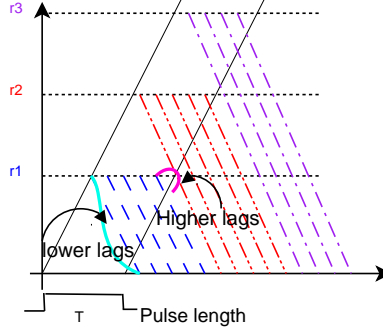


Figure 2.6 Range-time diagram in a long-pulse transmission. The fitting is performed on the data from altitudes r_1 , r_2 , and r_3 only. The number of altitudes contributing to the signal from r_1 decreases as we move to higher lags.

2.4.2 Height-by-height analysis

The height-by-height analysis has been traditionally used in the analysis of long-pulse modulation in which measurements suffer from range ambiguity. The method relies on the assumption that parameters do not vary over a range-gate (an altitude interval which, in the most common case, is equal to the distance covered by the transmitted pulse). One then fits the plasma parameters by least-squares methods to the chosen measured ACF using the theory of incoherent scatter at individual altitudes with range-gate separation (altitudes r_1 , r_2 , and r_3 in Figure 2.6 separated by range-gates).

Prior to least-squares estimation, however, a number of modifications should be applied to the received signal ACF. These changes are required to compensate for the range smearing, and without them the shape of the ACF becomes so distorted that the results of least-squares fitting become infeasible. One major reason is that the ACF of the received signal is the average of the ACF from several altitudes, and due to the nonlinear nature of the problem, the estimated parameters need not be close to the averages of the plasma parameters within the scattering volume.

Triangular weighting is a method which compensates the effect of range smearing and is based on weighing different lags of the measured ACF to make the effective weight of all lags equal. With a transmitted pulse of T s, the zeroth lag of the received signal ACF is the superposition of signals coming from T different altitudes separated by 1 s, whereas the $(T - 1)$ th lag measured signal ACF results from the signal coming from only one altitude. Thus, in effect, the range smearing weighs the zeroth lag T times higher than the $(T - 1)$ th lag, and the task of triangular weighting is to

undo this weighing process.

The hight-by-height analysis is also called range-gate analysis, as the data is processed at each range-gate separately. The technique, although simple and fast, suffers from the underlying unrealistic assumption along with the coarse resolution of estimated parameter profiles. Introduced biases in parameter profiles have also been reported in electron density in regions below the peak height [Lehtinen and Huuskonen, 1996; Holt et al., 1992].

2.4.3 Full-profile analysis

In this section, we first formalize a simple version of the idea of the full-profile analysis technique. We next explain the issues regarding practical implementations of such methods.

As opposed to the hight-by-height technique where only a few parameters are estimated using a nonlinear optimization procedure, in the full-profile method the unknown is a much longer vector giving the plasma variables at ionospheric-height grid points. This vector can be constructed from the elements of the following matrix

$$\mathbf{a} = \begin{bmatrix} T_e(r_1) & T_i(r_1) & N_e(r_1) & p(r_1) & \cdots \\ T_e(r_2) & T_i(r_2) & N_e(r_2) & p(r_2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_e(r_N) & T_i(r_N) & N_e(r_N) & p(r_N) & \cdots \end{bmatrix} \quad (2.20)$$

where r_i represent different grid points for $i = 1, \dots, N$. T_e , T_i , N_e and p denote the electron temperature, ion temperature, electron density, and composition, respectively. The method starts with initial estimates for all parameters at all altitudes. It then computes the theoretical ACF based on the parameters, imposes the range smearing (by taking into account the pulse shape), and makes a weighted comparison to the data, where weights are obtained from the data error covariances. The technique then proceeds by updating the parameter values so that the weighted difference between the actual data and the theoretical data is minimized.

The formulation above is in principle useful, but it is not the most efficient way to perform the full-profile analysis. The basic reason is that in order for the model to be accurate enough, the density of the range grid should be very high. In the best case, one is interested in a grid density

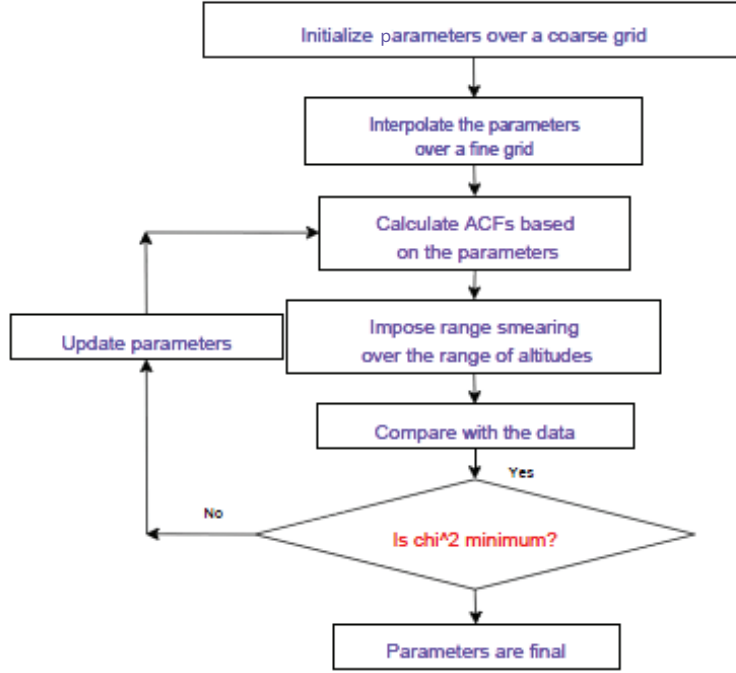


Figure 2.7 Flowchart of the effective implementation of the full-profile technique.

whose spacing is equal to the receiver sampling period. Unfortunately, considering such a dense grid increases the number of parameters of the optimization search space, which is rather costly in terms of computational power requirements, and it will be useful to develop more efficient ways for analysis.

Attempts have been made to develop effective implementation of the full-profile technique based on a hierarchy of grids and interpolation methods between the grids [Lehtinen *et al.*, 1996; Holt *et al.*, 1992]. The difference in interpolation methods makes the distinction among these various implementations. Holt *et al.* [1992] suggest Spline interpolation whereas Lehtinen *et al.* [1996] use Lagrange and linear interpolations throughout a hierarchy of grids. Figure 2.7 represents the flowchart of such methods.

Although performing the interpolation reduces the computational cost, the technique remains computationally expensive. One basic reason is due to the requirements of nonlinear computational techniques. One requirement is the computation of the derivative of the minimization function with respect to search variables (parameters in the coarse grid). Even though analytical expressions of

the derivatives of the lags of the theoretical auto-correlation function with respect to the ionospheric parameters are available, they cannot be exploited in the optimization procedures of the full-profile techniques. Therefore, forward difference, which slows down the speed of computation, is the only method that can be used for derivative calculation.

The full-profile technique is optimal in the sense that all the available information, such as the complete model of the ambiguity function and the full error covariance matrix, can be incorporated in the analysis. However, the computational cost of the method limits its routine implementation. In the next chapter, we develop the theory of a new hybrid inversion technique which aims at obtaining estimates that are close to optimal at a fraction of the usual computational cost. The technique is based on a correction to the effect of the transmitted waveform on the ACF lag profiles through a deconvolution process, and subsequent estimation of parameters from the plasma ACF at individual altitudes.

CHAPTER 3

PROPOSED TECHNIQUE FOR INVERSION OF INCOHERENT SCATTER MEASUREMENTS

The goal of this chapter is to introduce an efficient near-optimal technique for estimation of ionospheric parameters from incoherent scatter measurements. The technique is based on a correction to the effect of the transmitted waveform on the ACF lag profiles through a deconvolution process, and subsequent estimation of parameters from the plasma ACF at individual altitudes. In this work, we focus on long-pulse measurements and investigate the performance of the hybrid technique on simulated data (this chapter) and actual incoherent scatter radar measurements (Chapter 5).

In development of the hybrid technique, we exploit both the simplicity of height-by-height analysis and the accuracy of full-profile methods through considering the full model of ambiguity to present a simple, fast, and accurate method without the limitations of each of these common, currently used methods. For this purpose, we revisit the forward (direct) model of the incoherent scatter process as the form of the 1-dimensional convolution of the ionospheric ACF across range at each lag. We then present the matrix framework of convolution. The inversion technique is then formulated as the deconvolution of the lag profiles followed by the minimization of a least-squares cost function. Two different regularization methods for performing the deconvolution are discussed. Extension of the method to 2-dimensions is also discussed. The materials presented in Sections 3.1, 3.2, and 3.4.1 follow closely the description of the procedure reported by *Nikoukar et al.* [2008].

3.1 Forward model

This section describes the discretization of the radar equation (Chapter 2) and the corresponding matrix framework of the problem. Let $r' = \frac{ct}{2} - r$ and rewrite the radar equation as

$$\langle v(t)\overline{v(t+\tau)} \rangle \propto \int V_0(\frac{2r'}{c})\overline{V_0(\tau + \frac{2r'}{c})}R(\mathbf{k}, \frac{ct}{2} - r', \tau)dr' \quad (3.1)$$

where $\frac{ct}{2}$ refers to the altitude from which the signal is received (reference altitude). We discretize and approximate both sides of the above equation by a Reimann sum, as below

$$\langle v(\frac{2j\Delta r}{c})\overline{v(\frac{2j\Delta r}{c} + \tau)} \rangle \propto \sum_{i=-\infty}^{\infty} V_0(\frac{2i\Delta r}{c})\overline{V_0(\tau + \frac{2i\Delta r}{c})}R(\mathbf{k}, (j-i)\Delta r, \tau)\Delta r \quad \forall j \quad (3.2)$$

where $\Delta r = \frac{\text{sampling period}(\delta t) \times c}{2}$, and i and j are indexing terms. Note that i can range from $j - \frac{Tc}{2\Delta r} = j - \frac{T}{\delta t} = j - N_R$ to j , where N_R is the number of altitudes in each range-gate. This discretization is necessary since eventually the data and the resolution of the final parameter grid, Δr , are restricted by the sampling time of the receiver.

Following the notation introduced in Chapter 2, we replace $V_0(\frac{2i\Delta r}{c})\overline{V_0(\tau + \frac{2i\Delta r}{c})}$ in (3.2) by $\mathbf{p}_\tau(i)$ to obtain

$$\langle v(\frac{2j\Delta r}{c})\overline{v(\frac{2j\Delta r}{c} + \tau)} \rangle \propto \sum_{i=j-N_R}^j \mathbf{p}_\tau(i)R(\mathbf{k}, (j-i)\Delta r, \tau) \quad \forall j \quad (3.3)$$

Note that the above equation describes the relationship between the input and output of a linear time-invariant system, where the plasma ACF at certain time lag, τ , is the input, and the measured voltage ACF at the same time lag is the output. Moreover, the impulse response of the system is expressed as

$$\mathbf{p}_\tau(i) = V_0(\frac{2i\Delta r}{c})\overline{V_0(\tau + \frac{2i\Delta r}{c})} \quad \text{for } i = 0, 1, \dots, N_R \quad \text{and} \quad \tau = 0, \Delta\tau, \dots, T \quad (3.4)$$

where $\Delta\tau$ represents the time lag increment. Notice that $\mathbf{p}_\tau(i)$ preserves its form over different range-gates. Therefore, the index i can be considered to vary over a range-gate only, rather than being dependent on particular altitude j .

The above scheme can be visualized as incorporating the effects introduced by the transmitted pulse into a number of low-pass filters, each of which affects only one lag profile. These effects take place in the form of weighted averaging of the lag profiles, where the weights are determined by the ambiguity function at each time lag τ . Thus, the filter shape at each lag in the time domain is determined by the product of the pulse shape and its shifted version.

Once we envision range smearing as a filtering system, we can describe the relationship between its input and output, i.e. the plasma ACF at individual altitudes and the measured voltages ACF, as a convolution process; that is,

$$\langle v(\frac{2i\Delta r}{c})\overline{v(\frac{2i\Delta r}{c} + \tau)} \rangle = \mathbf{p}_\tau(i) * R(\mathbf{k}, i\Delta r, \tau) \quad \text{for } \forall i \text{ and } \tau = 0, \Delta\tau, \dots, T \quad (3.5)$$

where $*$ represents the convolution operation. With expansion of (3.5) with respect to all possible values of t and fixed value of τ , we can represent this convolution relationship in a matrix framework as

$$\mathbf{y}_\tau = \mathbf{P}_\tau \mathbf{m}_\tau \quad \text{for } \forall \tau \quad (3.6)$$

where

$$\mathbf{y}_\tau = \begin{bmatrix} v(\frac{i2\Delta r}{c})\overline{v(\frac{i2\Delta r}{c} + \tau)}|_{i=N_R} \\ v(\frac{i2\Delta r}{c})\overline{v(\frac{i2\Delta r}{c} + \tau)}|_{i=N_R+1} \\ \vdots \\ v(\frac{i2\Delta r}{c})\overline{v(\frac{i2\Delta r}{c} + \tau)}|_{i=n} \end{bmatrix}, \quad \mathbf{m}_\tau = \begin{bmatrix} R(\mathbf{k}, i\Delta r, \tau)|_{i=1} \\ R(\mathbf{k}, i\Delta r, \tau)|_{i=2} \\ \vdots \\ R(\mathbf{k}, i\Delta r, \tau)|_{i=n} \end{bmatrix}$$

$$\mathbf{P}_\tau = \begin{bmatrix} \mathbf{p}_\tau(i)|_{i=N_R} & \cdots & \mathbf{p}_\tau(i)|_{i=1} & \overbrace{\mathbf{0}}^{n-N_R} \\ 0 & \mathbf{p}_\tau(i)|_{i=N_R} & \cdots & \mathbf{p}_\tau(i)|_{i=1} & \overbrace{\mathbf{0}}^{n-(N_R+1)} \\ & & \vdots & & \\ \overbrace{\mathbf{0}}^{n-N_R} & & \mathbf{p}_\tau(i)|_{i=N_R} & \cdots & \mathbf{p}_\tau(i)|_{i=1} \end{bmatrix}$$

Notice that \mathbf{y}_τ is the vector of measurement ACF at time lag τ from all altitudes (lag profiles). Similarly, \mathbf{m}_τ includes the altitude profile (for n altitudes) of the true plasma ACF at the same time lag. The convolution matrix \mathbf{P}_τ relates the measurement ACF and plasma ACF at lag τ .

A more complete model of ISR measurement can be rewritten as

$$\mathbf{y}_\tau = \mathbf{P}_\tau \mathbf{m}_\tau + \epsilon_\tau \quad \forall \tau \quad (3.7)$$

where ϵ_τ represents the measurement error at time lag τ . In general this error is signal-dependent especially when the backscattered signals are strong due to high electron densities or high transmitted power. Because of this dependence, it is not possible to obtain an estimate of the lag profiles using only one single transmission. Instead, the measured ACFs should be added for several pulse transmissions to improve statistical accuracy (integration). The data is typically integrated for several seconds. Although signal-dependent, the noise gains Gaussian characteristics as signal is integrated over many pulses (according to the central limit theorem). Therefore, it can be well described by its mean, $\mu = \langle \epsilon_\tau \rangle$, and covariance matrix, $\Sigma_{\epsilon_\tau} = \langle \epsilon_\tau \epsilon_\tau^T \rangle$. The covariance matrix will be diagonal if the errors in different lags are independent. Otherwise it will have nonzero off-diagonal elements.

3.2 Inverse model

In the previous sections we established the relationship between the plasma ACF at individual altitudes and the received voltage ACF as a convolution process, where the shape of the convolving function is dependent on the pulse envelope as well as the time lag values. We exploit this property in our proposed inversion method, a detailed description of which is presented in the following sections.

3.2.1 Deconvolution

The major motivation for the deconvolution of the lag profiles is to remove the range smearing from the measured signal ACF and obtain the plasma ACF at single altitudes. The elimination of range ambiguity allows us to use analytical derivatives of the theoretical ACF lags with respect to ionospheric parameters, as opposed to forward differences, in least-squares optimization algorithms and, hence, to reduce the computational cost significantly. Furthermore, when using deconvolution methods, one does not require the imposition of unrealistic assumptions on the parameter profiles,

such as stationarity, for the whole range-gate as is the case in height-by-height analysis.

Regularization

To perform the deconvolution task, we can use methods such as inverse filtering and least-squares analysis [Blahut, 2004; Lagendijk and Biemond, 2000; Karl, 2000]. These methods, although straightforward and easy to implement, suffer from a common drawback, which is the instability of the solution in the face of perturbations to data. This major drawback raises the need for regularization. Through regularization, we impose *a priori* knowledge about the underlying process to stabilize the solution in the presence of noise and to permit the identification of physically reasonable estimates of parameters of interest. A regularization method can be considered as a modified least-squares technique, where the modifications appear in the form of additional constraints to the residual norm defined in (3.8) as side constraint norms. More precisely, we can represent the regularized estimate as the solution to the following minimization problem

$$\hat{\mathbf{m}}_{\tau,reg} = \arg \min_{\mathbf{m}_{\tau}} \left(\|\mathbf{y}_{\tau} - \mathbf{P}_{\tau} \mathbf{m}_{\tau}\|_{\Sigma_{\epsilon_{\tau}}^{-1}}^2 + \sum_i \lambda_i C_i(\mathbf{m}_{\tau}) \right) \quad \forall \tau \quad (3.8)$$

where λ_i and C_i are the i th regularization parameter and regularization functional, respectively. The first term controls data fidelity (i.e. how closely the solution fits the data), whereas the second term (the regularization term) controls how well the solution matches our prior knowledge. The role of the regularization parameter can be viewed as controlling the trade-off between the impact of data and the impact of *a priori* knowledge on the solution.

In what follows we introduce two methods of regularization which use different side functionals.

A: Tikhonov regularization The most common regularization method is the Tikhonov regularization with a quadratic functional [Karl, 2000; Demoment, 1989]. The general expression for the Tikhonov method is

$$\hat{\mathbf{m}}_{\tau,Tikhonov} = \arg \min_{\mathbf{m}_{\tau}} \left(\|\mathbf{y}_{\tau} - \mathbf{P}_{\tau} \mathbf{m}_{\tau}\|_{\Sigma_{\epsilon_{\tau}}^{-1}}^2 + \lambda^2 \|\mathbf{L} \mathbf{m}_{\tau}\|^2 \right) \quad \forall \tau \quad (3.9)$$

where the most common choice for \mathbf{L} is the discrete representation of the gradient operator which

enforces a roughness penalty and, hence, a smoothness constraint. As an example, the discretized first-order gradient operator can be represented as

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & \cdots & \\ 0 & 1 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 1 & -1 & \end{bmatrix} \quad (3.10)$$

Note that $\|\mathbf{z}\|_p^p = (\sum_i z_i^p)^{1/p}$. In this case $\|\mathbf{L}\mathbf{m}_\tau\|^2$ is a measure of the variability of the estimate. Therefore, the overall functionality of the method can be visualized as penalizing large gradients of the plasma ACF lag profiles, resulting in smoother lag profiles where the degree of smoothness depends on the value of the regularization parameter.

The solution to the minimization in (3.8), i.e. the Tikhonov regularized estimate, can be obtained as the solution to the following set of equations

$$(\mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{P}_\tau + \lambda^2 \mathbf{L}^T \mathbf{L}) \hat{\mathbf{m}}_\tau = \mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{y}_\tau \quad \forall \tau \quad (3.11)$$

B: Total variation (TV) regularization One drawback associated with Tikhonov regularization is that it severely penalizes the sharp gradients in ACF lag profiles; therefore, if there is a natural sharp gradient in the electron density profile, for example, it will not be recovered by the Tikhonov method.

Total variation regularization is a nonlinear technique that tries to preserve sharp gradients in profiles [Karl, 2000; Vogel and Oman, 1996]. The general expression for the TV method is

$$\hat{\mathbf{m}}_{\tau,TV} = \arg \min_{\mathbf{m}_\tau} \left(\|\mathbf{y}_\tau - \mathbf{P}_\tau \mathbf{m}_\tau\|_{\Sigma_{\epsilon_\tau}^{-1}}^2 + \|\mathbf{L}\mathbf{m}_\tau\|_1 \right) \quad \forall \tau \quad (3.12)$$

The ℓ_1 norm used in the TV technique does not penalize the sharp edges in lag profiles as severely as the quadratic norm used in the Tikhonov method. Thus it is well suited for situations where the plasma parameter altitude profiles contain sharp gradients.

The difficulty with the formulation in (3.12) is that it is not differentiable around the origin,

and thus it leads to a challenging nonlinear optimization problem. One approach to overcome this challenge is to consider the TV estimate as the solution to the following sets of equations in the limit as $\beta \rightarrow 0$ [Karl, 2000]

$$(\mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{P}_\tau + \lambda^2 \mathbf{L}^T \mathbf{W}_\beta(\hat{\mathbf{m}}_{\tau,TV}) \mathbf{L}) \hat{\mathbf{m}}_{\tau,TV} = \mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{y}_\tau \quad \forall \tau \quad (3.13)$$

where the diagonal weight matrix $\mathbf{W}_\beta(\hat{\mathbf{m}})$ depends on \mathbf{m} , and β and is given by

$$\mathbf{W}_\beta(\hat{\mathbf{m}}) = \frac{1}{2} \text{diag} \left[\frac{1}{\sum_i \sqrt{[\mathbf{L}\mathbf{m}]_i^2 + \beta}} \right] \quad (3.14)$$

with $\beta > 0$ a constant. Equation (3.13) is obtained by smoothly approximating the ℓ_1 norm of the derivative: $\|\mathbf{L}\mathbf{m}\|_1 = \sum_i \sqrt{[\mathbf{L}\mathbf{m}]_i^2 + \beta}$.

Computationally, (3.13) is still nonlinear, since the weight matrix depends on \mathbf{m} . However, it suggests a simple fixed-point iteration for the TV estimate $\mathbf{m}_{\tau,TV}$, only requiring the solution of the standard linear problem at each step

$$(\mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{P}_\tau + \lambda^2 \mathbf{L}^T \mathbf{W}_\beta(\hat{\mathbf{m}}_\tau^{(k)}) \mathbf{L}) \hat{\mathbf{m}}_\tau^{(k+1)} = \mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{y}_\tau \quad \forall \tau \quad (3.15)$$

where β is typically set to a small value.

Error estimation in regularized solutions

Although the above deterministic view of regularization provides an expression for the estimate, it lacks the ability to provide measures of the solution uncertainty. To obtain the estimation errors, we need to cast the inversion problem into a statistical framework by choosing an appropriate prior statistical model for \mathbf{m}_τ and seeking a maximum *a posteriori* probability (MAP) estimate [Demoment, 1989; Kamalabadi et al., 1999].

In order to apply the MAP estimation to our inversion problem (Equation (4.12)), we first need to view the noise, ϵ_τ , and the unknown lag profile, \mathbf{m}_τ , as random vectors. Then the MAP estimate of \mathbf{m}_τ can be expressed as the vector which maximizes the *a posteriori* density $p(\mathbf{m}_\tau | \mathbf{y}_\tau)$. Using

the Bayes rule and the monotonicity properties of the logarithm, we obtain

$$\hat{\mathbf{m}}_\tau = \arg \max_{\mathbf{m}_\tau} p(\mathbf{m}|\mathbf{y}_\tau) = \arg \max_{\mathbf{m}_\tau} \ln p(\mathbf{y}_\tau|\mathbf{m}_\tau) + \ln p(\mathbf{m}_\tau) \quad (3.16)$$

where $\ln(\cdot)$ represents the natural logarithm. Notice that, similarly to the deterministic optimization objective function (3.8), the above optimization problem consists of two terms: the first term is data dependent and is called the log-likelihood function, while the second term is dependent only on \mathbf{m}_τ and is called the prior model.

Let us now consider the case of Gaussian statistics for noise (see Section 3.1) and prior model

$$\mathbf{y}_\tau = \mathbf{P}_\tau \mathbf{m}_\tau + \epsilon_\tau \quad \forall \tau \quad (3.17)$$

$$\epsilon_\tau \sim \mathcal{N}(\mathbf{0}, \Sigma_{\epsilon_\tau}) \quad \text{and} \quad \mathbf{m}_\tau \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{m}_\tau}) \quad (3.18)$$

where $\mathbf{m}_\tau \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{m}_\tau})$ denotes that \mathbf{m}_τ is a Gaussian distributed random vector with mean $\mathbf{0}$ and covariance $\Sigma_{\mathbf{m}_\tau}$.

We can now provide an expression for the associated measure of uncertainties, using the statistical framework and through the error covariance matrix $\Sigma_{\hat{\mathbf{m}}_\tau} = \langle \mathbf{e}_\tau \mathbf{e}_\tau^T \rangle$, where $\mathbf{e}_\tau = \mathbf{m}_\tau - \hat{\mathbf{m}}_\tau$. It has been shown that for the MAP estimate in (3.16), the error covariance is given by [Demoment, 1989]

$$\Sigma_{\hat{\mathbf{m}}_\tau} = (\mathbf{P}_\tau^T \Sigma_{\epsilon_\tau}^{-1} \mathbf{P}_\tau + \Sigma_{\mathbf{m}_\tau}^{-1})^{-1} \quad (3.19)$$

For Tikhonov regularization, under the Gaussian assumptions and upon substitution in (3.16), we obtain

$$\hat{\mathbf{m}}_\tau = \arg \min_{\mathbf{m}_\tau} \left(\|\mathbf{y}_\tau - \mathbf{P}_\tau \mathbf{m}_\tau\|_{\Sigma_{\epsilon_\tau}^{-1}}^2 + \|\mathbf{m}_\tau\|_{\Sigma_{\mathbf{m}_\tau}^{-1}} \right) \quad (3.20)$$

Comparison of Equations (3.9) and (3.20) yields that they are equivalent under certain conditions. This equality condition is only related to the prior model, since the data fidelity terms are equal.

The appropriate prior model corresponds to

$$\mathbf{L}\mathbf{m}_\tau = \mathbf{r}_\tau, \quad \mathbf{r}_\tau \sim \mathcal{N}(0, \lambda^2 \mathbf{I}) \quad \forall \tau \quad (3.21)$$

where \mathbf{L} is a discrete approximation of the gradient operator. According to this prior model, \mathbf{m}_τ itself is Gaussian distributed with covariance $\Sigma_{\mathbf{m}_\tau} = \lambda^2 (\mathbf{L}^T \mathbf{L})^{-1}$. In addition, the increments of \mathbf{m}_τ are uncorrelated with variance $\lambda^2 \mathbf{I}$; thus \mathbf{m}_τ itself corresponds to a Brownian motion-type model.

For TV regularization, by comparing Equations (3.12) and (3.16), we find it reasonable to make the following probabilistic association for the prior

$$\ln p(\mathbf{m}) \propto -\lambda^2 \sum_i |[\mathbf{L}\mathbf{m}]_i| p(\mathbf{m}) \sim \prod_i e^{-\lambda^2 |[\mathbf{L}\mathbf{m}]_i|} \quad (3.22)$$

The statistical prior model for \mathbf{m} has increments $[\mathbf{L}\mathbf{m}]_i$ that are independently identically distributed (IID) according to Laplacian density.

Notice that the diagonal entries of $\Sigma_{\hat{\mathbf{m}}_\tau}$ are the variances of individual estimation errors. This error covariance matrix is used as an uncertainty level of the deconvolved lag profiles in the next step of our proposed method, which is the nonlinear least-squares optimization procedure. We still need to consider the choice of the regularization parameter, as it has a direct effect on the solution.

Choice of regularization parameter

Regularization inherently involves the trade-off between fidelity to the data, as measured by the residual norm, and fidelity to *a priori* information, as measured by the side constraint norm. By using a gradient operator as the regularization functional, the overall functionality of the method can be visualized as penalizing large gradients of the solution, which results in smoother lag profiles, and the degree of smoothness depends on the value of regularization parameter. Therefore, an important part of the regularization technique is to find a reasonable value for λ . From various methods of choosing λ [Karl, 2000], in this work we focus on the L-curve and unbiased predictive risk estimator (UPRE) methods.

A: L-curve The L-curve method does not require knowledge of the noise characteristics. An L-curve is a plot of the side constraint norm versus the residual norm as λ varies and has a characteristic L-shape. In general, it consists of vertical and horizontal parts, where the vertical part corresponds to unregularized estimates (where the solution is dominated by the amplifying noise, and small changes in λ have a large effect on the solution), and the horizontal part corresponds to overregularized solutions. In this region, changes to λ affect the solution weakly, but produce large changes in the residual error. The regularization parameter is chosen to be the corner between the horizontal and vertical portions of the curve, which defines the transition between the over- and under-regularization and thus represents the balance between these two extremes.

B: Unbiased predictive risk estimator (UPRE) The UPRE method is another technique for regularization selection, and unlike the L-curve method, it relies on a good estimate of the measurements error covariance matrix. The UPRE method finds the regularization parameter that minimizes the predictive risk, defined as [Vogel, 2002]

$$L = \|\mathbf{p}_\tau\|_2^2 = \|\mathbf{y}_\tau - \mathbf{y}_\tau^{true}\|_2^2 \quad \forall \tau \quad (3.23)$$

where $\mathbf{y}_\tau^{true} = \mathbf{P}_\tau \mathbf{m}_\tau^{true}$. Although \mathbf{y}_τ^{true} cannot be directly computed, it can be estimated based on properties of the measurement noise. The UPRE estimator can be represented as

$$U(L) = \|\hat{\mathbf{y}}_\tau - \mathbf{y}_\tau\|_{\Sigma_\epsilon}^2 + 2\text{Tr}(\mathbf{M}_\lambda) \quad (3.24)$$

where \mathbf{M}_λ is the so-called “influence matrix,” and provides the link between the predictive data, $\hat{\mathbf{y}}_\tau$, and actual measurements, \mathbf{y}_τ , i.e. $\hat{\mathbf{y}}_\tau = \mathbf{M}_\lambda \mathbf{y}_\tau$. For the Tikhonov regularization, \mathbf{M}_λ can be expressed as

$$\mathbf{M}_\lambda = \Sigma_\epsilon^{-1/2} \mathbf{P}_\tau^T (\mathbf{P}_\tau^T \Sigma_\epsilon^{-1} \mathbf{P}_\tau + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{P}_\tau \Sigma_\epsilon^{-1/2} \quad (3.25)$$

3.2.2 Nonlinear least-squares (NLLS) optimization

Once the range smearing effect of the transmitted waveform is removed via deconvolution, we perform LS optimization to estimate ionospheric parameters. We minimize the following objective function for each altitude separately

$$\chi^2 = (\mathbf{m} - g(\mathbf{a}))^T \Sigma_{\hat{\mathbf{m}}}^{-1} (\mathbf{m} - g(\mathbf{a})) \quad (3.26)$$

where \mathbf{m} is the deconvolved (plasma) ACF at a certain altitude and is obtained from \mathbf{m}_τ for all values of τ , \mathbf{a} represents the parameter vector $(T_e, T_i, N_e, p, \dots)$, and g is the nonlinear incoherent scatter function which relates the ionospheric parameters to the plasma spectrum, or equivalently its auto-correlation function. Moreover, $\Sigma_{\hat{\mathbf{m}}}$ is the error covariance matrix of the deconvolved ACF and is different from that of the measured voltage lag profiles (convolved ACF) in (4.12). In any nonlinear least-squares estimation, the errors play an important role since they provide information about the reliability of each data point. In the case of independent and equal errors, the error covariance matrix can be written as $\Sigma_{\hat{\mathbf{m}}} = \sigma^2 I$, where σ^2 denotes the error, and (3.26) can be simplified as

$$\chi^2 = \|\mathbf{m} - g(\mathbf{a})\|_{\sigma^2 I}^2 \quad (3.27)$$

The situation, however, becomes more complicated when the errors are correlated, as is the case for plasma ACF measurements. In order to make (3.26) as simple as (3.27) in the general case of dependent errors, we need to factor out the inverse of the deconvolved ACF error covariance matrix ($\Sigma_{\hat{\mathbf{m}}}$) by Cholesky decomposition into matrices \mathbf{D} and \mathbf{D}^T such that $\Sigma_{\hat{\mathbf{m}}}^{-1} = \mathbf{D}^T \mathbf{D}$. With this factorization, (3.26) can be rewritten as

$$\begin{aligned} \chi^2 &= (\mathbf{m} - g(\mathbf{a}))^T \mathbf{D}^T \mathbf{D} (\mathbf{m} - g(\mathbf{a})) \\ &= (\mathbf{D}(\mathbf{m} - g(\mathbf{a})))^T (\mathbf{D}(\mathbf{m} - g(\mathbf{a}))) \\ &= \|\mathbf{D}(\mathbf{m} - g(\mathbf{a}))\|_I^2 \end{aligned} \quad (3.28)$$

which has a similar structure to that of (3.27) and hence is easy to implement.

3.3 Extension to 2-dimensions

In derivation of the above-mentioned hybrid methodology, it has been assumed that the sampling of the receiver is instantaneous. In other words, the receiver impulse response is approximated as a Dirac delta function; as such, the incoherent scatter radar function, which the forward model is built upon, is considered as a function of range only. In most practical circumstances, however, the receiver sampling is not instantaneous and occurs only after some filtering. This essentially means that in addition to being averages of the plasma fluctuation correlations over a certain range interval, the measured lag products are also averages of plasma correlations over a certain delay interval. Developing a hybrid technique for this situation demands consideration of the full 2-dimensional incoherent scatter radar function.

In order to develop the full 2-dimensional ambiguity function, we need to revisit (2.3) which represents the scattered signal from the electron density fluctuations in the ionosphere. One should note that the reception of the signal happens only after some filtering in time. The more complete version of this equation can be represented by

$$v(t) \propto \int du h(t-u) \int_{\mathbf{r}} n(\mathbf{r}, u) e^{-j2k_0 r} V_0(u - \frac{2r}{c}) d\mathbf{r} \quad (3.29)$$

where $h(t)$ is the impulse response of the receiver. Note that if $h(t) = \delta(t)$ the above expression reduces to (2.3). By computing the ACF of the received signal, we obtain

$$\begin{aligned} \langle v(t) \overline{v(t+\tau)} \rangle &\propto \int_{u'} du' \int_u du \int_{\mathbf{r}'} d\mathbf{r}' \int_{\mathbf{r}} d\mathbf{r} \langle n(\mathbf{r}, u) \overline{n(\mathbf{r}', u' + \tau)} \rangle h(t-u) V_0(u - \frac{2r}{c}) \\ &\quad \overline{h(t+\tau-u') V_0(u' - \frac{2r'}{c})} e^{-2jk_0(r'-r)} \end{aligned} \quad (3.30)$$

Let $u' = u + l$, and using the approximations of Section 2.1.1, rewrite the above equation as

[*Lehtinen and Huuskonen, 1996; Woodman, 1991*]

$$\langle v(t)\overline{v(t+\tau)} \rangle \propto \int_{\mathbf{r}} d\mathbf{r} \int_u du h(t-u)V_0(u-\frac{2r}{c}) \quad (3.31)$$

$$\frac{\int_l dl h(t+\tau-u-l)V_0(u+l-\frac{2r'}{c})R(\mathbf{k}, \mathbf{r}, l)}{\int_{\mathbf{r}} d\mathbf{r} \int_l dl \int_u du h(t-u)V_0(u-\frac{2r}{c})h(t+\tau-u-l)V_0(u+l-\frac{2r}{c})R(\mathbf{k}, \mathbf{r}, l)} \quad (3.32)$$

By inspecting (3.33) we realize that the ACF of the received voltage at time lag τ can be represented as a weighted average of the plasma ACF in both range and lag directions. The weight is dependent on both the impulse response of the receiver and the envelope of the transmitted pulse, and can be expressed as

$$w_{\tau}^t(l, \frac{2r}{c}) = \int du h(t-u)V_0(u-\frac{2r}{c})\overline{h(t+l-u+\tau)V_0(u-l-\frac{2r}{c})} \quad (3.33)$$

This averaging operation with respect to the delay (l) and range (r) variables is expressed by $w_{\tau}^t(l, \frac{2r}{c})$ for a given time t and time lag τ in the following generalized incoherent scatter radar equation [*Lehtinen and Huuskonen, 1996; Woodman, 1991*]

$$\langle v(t)\overline{v(t+\tau)} \rangle = \int_r d\mathbf{r} \int_l dl w_{\tau}^t(l, \frac{2r}{c})R(2\mathbf{k}, \mathbf{r}, l) \quad (3.34)$$

Similarly to the 1-dimensional case, casting this 2-dimensional integral equation into a matrix framework allows for utilization of advanced efficient inversion techniques. The fundamentals of the hybrid technique developed in Section 3.2 are amenable to this problem, as well. One can form the data vector \mathbf{y} and the unknown vector \mathbf{m} , consisting of measured signal ACF at a particular time lag and plasma ACF at several delays for several heights, respectively, as well as the forward model, $w_{\tau}^t(l, \frac{2r}{c})$, to yield the following linear formulation: $\mathbf{y} = \mathbf{W}_{\tau}\mathbf{m} + \epsilon$, where ϵ represents the noise. An illustration of the full forward model is given in Figure 3.1 (a) for a transmitted pulse of length 300 μs and a receiver boxcar function of length 20 μs . This figure includes the ambiguity functions for several lags altogether.

Note that when the variation of the plasma ACF is much slower than the receiver impulse response

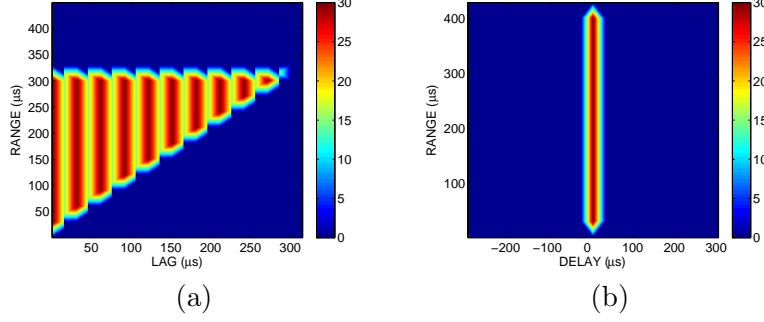


Figure 3.1 (a) Full 2-dimensional ambiguity function for a transmitted pulse of length $300 \mu\text{s}$ and a receiver boxcar function of length $20 \mu\text{s}$. (b) Ambiguity function for lag 0.

length, one can ignore the variation of the ambiguity function in the delay direction, and as such the consideration of the full 2-dimensional model is not necessary. Because this condition is met for Arecibo F-region measurements, we use the 1-dimension approximation presented in Section 2.1.1 for data inversion (See Chapter 5).

3.4 Numerical results

In this section, we present several numerical results. In Section 3.4.1 we first describe a method to simulate the noise-like behavior of the incoherent scatter signal. We then compare the performances of the hybrid technique with the conventional processing techniques. In Section 3.4.2 we compare the performance of the Tikhonov and TV regularization techniques in situations where the plasma altitude profiles are not smooth.

3.4.1 Hybrid technique versus conventional processing techniques

In this section we investigate the performance of the hybrid technique on a numerical experiment and compare the corresponding estimation results with those of common methods currently used in terms of mean-squared error and computation time.

In data simulation, we take into account the fluctuations of both the incoherent scatter signal and the receiver noise by simulating the received signals. The data simulation algorithm is summarized as follows:

1. Generate the altitude profiles of the ionospheric parameters. For simplicity, we restrict the ions in the ionosphere to oxygen only, which is a realistic assumption in the lower F-region and around the height of peak electron density. Figure 3.2 shows an example of three parameter profiles (profiles of electron and ion temperatures and electron density) used in the simulations.
2. Generate noisy voltage samples based on the parameters at each altitude and modulated pulse shape, which itself requires the following steps:
 - (a) Form the incoherent scatter spectrum at individual altitudes based on parameter profiles.
 - (b) Generate two sets of random spectra, each based on the square root of the spectrum at each altitude. With this configuration, the generated signal will have stronger variations whenever the spectrum itself is stronger (consideration of signal fluctuation).
 - (c) Form the complex spectra.
 - (d) Transform into the time domain.
 - (e) Multiply the complex ACF by the pulse shape.
3. Add voltage samples from different heights.
4. Add random noise to the voltage samples generated in (3). (This random noise represents sky and receiver noise and is independent from the signal.)
5. Form lag-profiles.
6. Accumulate over many pulse transmissions.

Note that a radar frequency of 430 MHz and an un-coded long pulse of length $280 \mu\text{s}$ are used in simulations. The lag spacing is set to $10 \mu\text{s}$, and the data is accumulated over 12000 pulse transmissions, which yields an error of about 4% in the lags of the measured ACF.

In the implementation of the hybrid technique, we apply Tikhonov regularization with the first-order gradient operator (refer to (3.10)) on the lag profiles in the first step. The deconvolved lag profiles are then obtained by solving the set of equations denoted in (3.11) for each time lag (with integer multiple of lag spacing $(\Delta\tau)$) individually. From various approaches that exist to solve

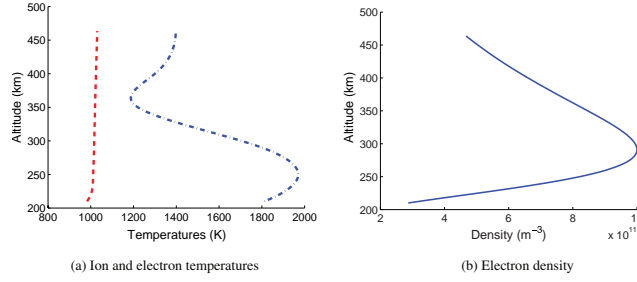


Figure 3.2 Profiles of a hypothetical ionosphere used in the simulations. For simplicity, this ionosphere is comprised of a single ion species. This is a realistic assumption in the lower F-region and near the height of peak electron density.

the equations in (3.11), we use the conjugate gradient method, which is a fast converging iterative method [Golub and Van Loan, 1989]. In the second step of the method, the Levenberg-Marquardt algorithm [Levenberg, 1944; Marquardt, 1963] is used as the NLLS optimization procedure.

The proper value of the regularization parameter, λ , is achieved by the means of the L-curve shown in Figure 3.3 along with the proper regularization parameter ($\lambda \simeq 25.6$), identified by the dashed lines. Figure 3.4 illustrates the estimated parameter profiles for this regularization parameter. While using smaller values of this parameter results in a much larger variation in the estimated parameters, using larger values introduces biases into the estimation results (see Figure 3.5). The ionospheric parameter estimation errors are also plotted in Figure 3.6(a)-(c) as a function of regularization parameters, where the errors are measured as $\sqrt{\frac{\sum_{j=1}^n (a_{ij} - \hat{a}_{ij})^2}{n}}$ with a_{ij} and \hat{a}_{ij} denoting the i th ideal and estimated ionospheric parameter at altitude j , respectively. According to this figure, there is a range of regularization parameters over which the parameter estimation errors are minimum (or within 10% of the minimum), and the proper regularization parameters selected by the L-curve fits well in this range. Figure 3.6(d) also illustrates the computation time required to complete the analysis on a 2.4 GHz workstation.

The height-by-height estimation results, which are obtained by the application of triangular weighting prior to LS optimization, are given in Figure 3.7. Note that the fitting is performed only at parameter grids specified by squares, and the estimated parameter profiles (dotted curves) are produced by linear interpolations on these grids. The coarse resolution achieved by this method is evident from the plots.

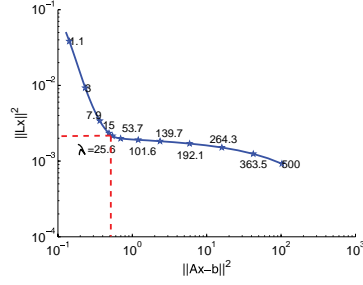


Figure 3.3 L-curve for Tikhonov regularization applied in the first step of the proposed method. As the regularization parameter increases, the curve moves from the upper left to the lower right. The best choice of the regularization parameter corresponds to the point of maximum curvature (corner) of the L-curve. This point ($\lambda = 25.6$) is identified by the dashed lines.

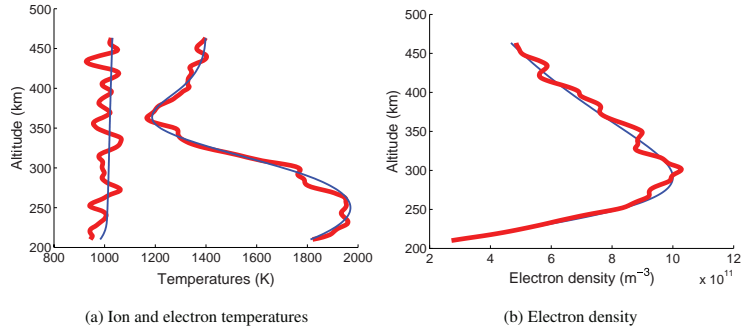


Figure 3.4 The proposed hybrid method estimation results (thick curves). Original profiles are represented by the thin lines. The regularization parameter has been chosen based on the corner point in Figure 3.3.

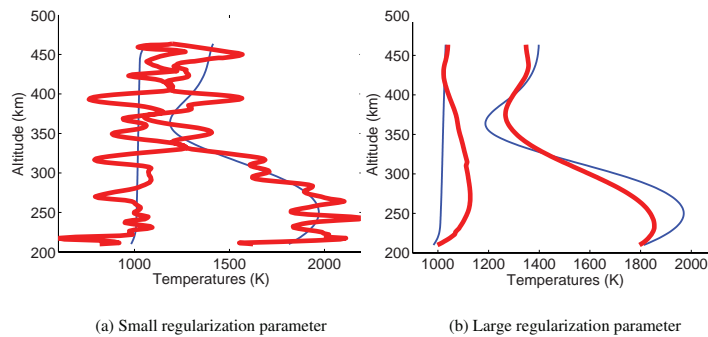


Figure 3.5 Estimated ion and electron temperature profiles for small and large regularization parameters (thick curves). The thin curves show the original profiles. While using a very small value of this parameter results in a much larger variation in the estimated parameters, using a large value introduces biases into the estimation.

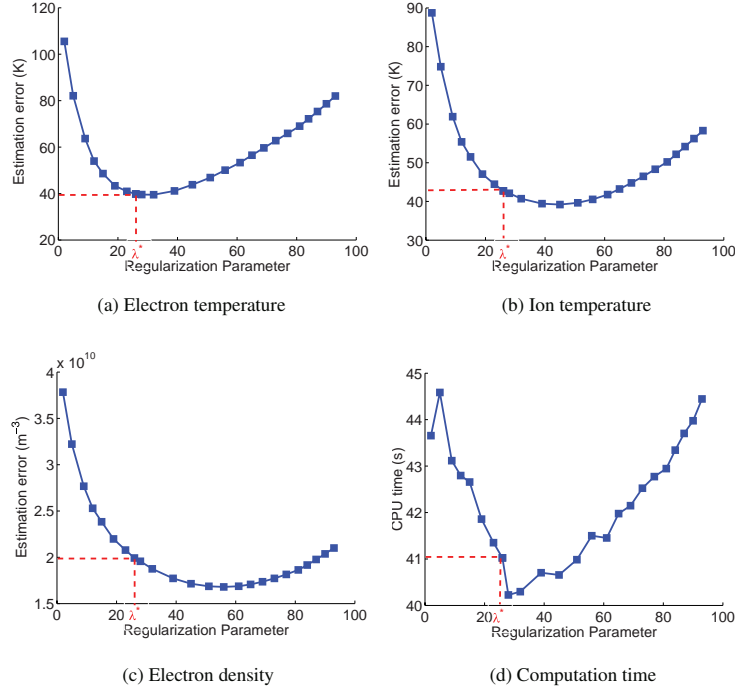


Figure 3.6 Performance of the hybrid technique (in terms of estimation error and computation time) as regularization parameter varies. The estimation errors for the regularization parameter selected by the L-curve is denoted by dashed lines.

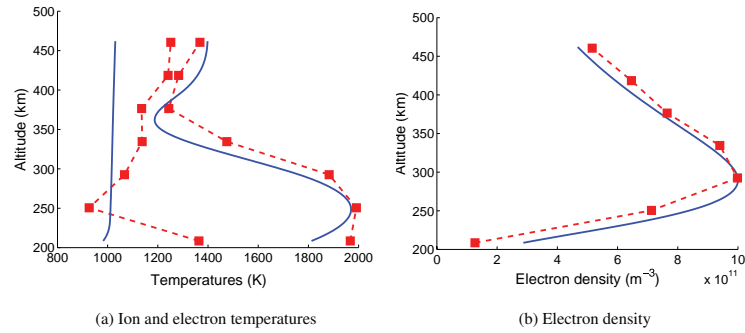


Figure 3.7 Height-by-height analysis estimation results. Note that the LS fitting is performed only at parameter grids specified by squares and the estimated parameter profiles (dotted curves) are produced by linear interpolations on these grids. The original profiles are given by solid lines.

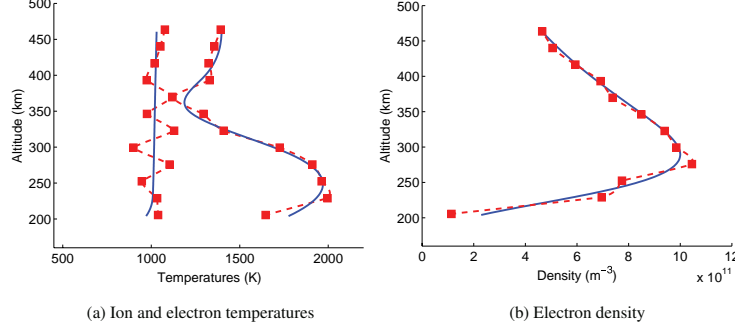


Figure 3.8 Optimal full-profile estimated parameters obtained for a parameter grid with 26 km spatial resolution (dotted curves). Original profiles are represented by solid lines.

The optimal parameter estimation results of the full-profile technique are shown in Figure 3.8. The resolution of the parameter grid in this optimal case is 26 km. The optimality is determined based on the performance of the technique over a wide range of parameter grid resolutions. The plots in Figure 3.9(a)-(d) show the errors in estimated parameters as well as the timing required to complete the analysis (on a 2.4 GHz workstation) for parameter grids with spatial resolutions from 10 to 60 km (corresponding to 30– to 6–point parameter grids). These errors are minimum (or very close to minimum) over a limited range of spatial grid resolutions. This range of optimal grid densities varies slightly for different ionospheric parameters. But in general we can consider the interval between 21 and 29 km to be within 10% from the optimal. For comparison, we also plot the estimated temperatures for the two spatial resolutions of 57 km and 12 km in Figure 3.10. The high variation of the estimated parameters from the true parameters in both cases, due to under-fitting or over-fitting of the solution to the noisy data, is evident from the plots.

From Figures 3.6 (a)-(c), 3.7, and 3.9 (a)-(c), it is clear that the hybrid technique is able to provide faithful estimates of the ionospheric parameters that are nearly as good as the ones given by the full-profile technique (provided that the optimal regularization parameter and grid density are chosen) while it outperforms the height-by-height method in terms of estimation accuracy. Two main differences between the performances of the two methods, however, can be inferred from the plots in Figures 3.6(d) and 3.9(d). First, the hybrid method requires nearly constant computational time for various regularization parameters, unlike the full-profile technique where the analysis time is linearly dependent on the parameter grid density. Second, the hybrid technique

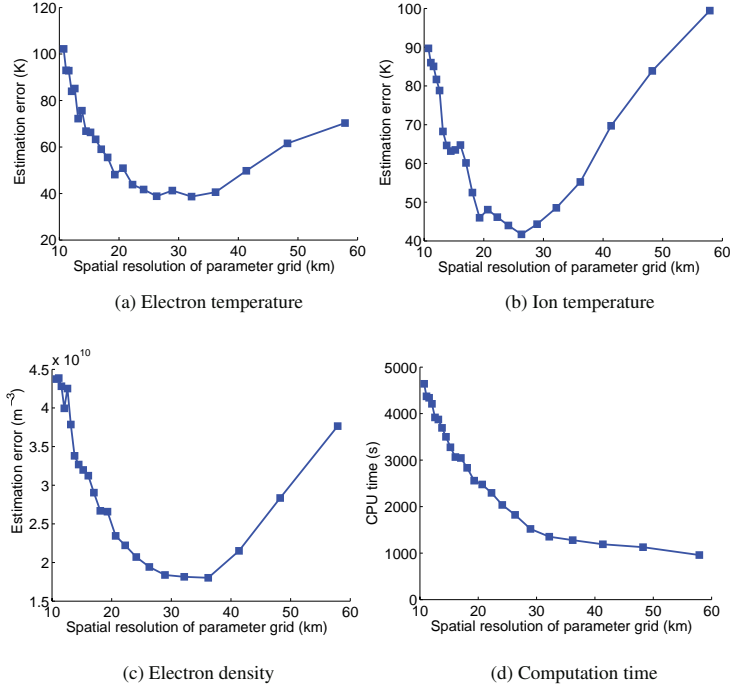


Figure 3.9 Performance of the full-profile technique (in terms of estimation error and computation time) as a function of parameter grid resolution.

reduces the computation time to nearly 95% of that of the full-profile technique. The restriction of the optimization search space to the number of ionospheric parameters at each range and hence the ability to utilize analytical derivatives of the ACF lags with respect to the physical parameters serves as the main key to the efficiency of the method.

Note that in our simulation setup, the errors in lag estimates become highly correlated. However, we include only the variances of the errors (diagonal elements of the error covariance matrix) in the above analysis because the computation of the full covariance matrix is not tractable, especially as the number of altitudes increases. The estimation errors noted in Figures 3.6 and 3.9(a)-(c) have been obtained numerically using Monte Carlo-type simulations. Therefore, no underestimation of errors occurs as is the case when errors are obtained analytically using the diagonal error covariance matrix [Huuskonen and Lehtinen, 1996]. Nevertheless, the high correlations in the parameter profiles estimated by the hybrid method are in part the consequences of ignoring these error covariances.

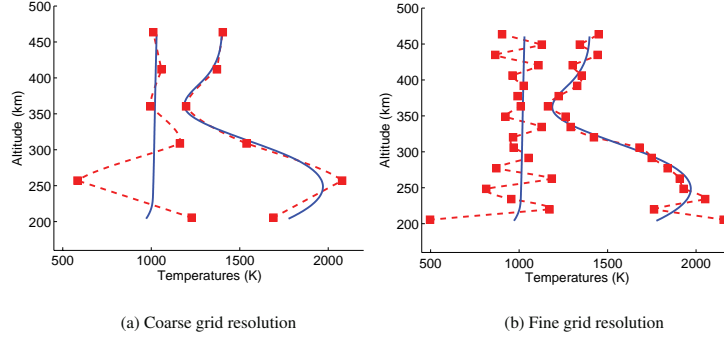


Figure 3.10 Full-profile estimated temperatures for two different grid parameters (spatial resolution of 57 km in (a) and 12 km in (b)). The original profiles are represented by solid lines. The high variation of the estimated parameters from the true parameters in both cases due to under-fitting or over-fitting of the solution to the noisy data is evident from the plots.

3.4.2 Quadratic versus edge-preserving regularization techniques

In this section we compare the performances of the Tikhonov and TV regularization techniques. In Section 3.2.1 it was mentioned that the Tikhonov method results in over-smoothness of the ACF lag profiles. Therefore, in cases where the parameter altitude profiles are not smooth, one has to use edge-preserving methods such as the TV technique. As an example of a sharp gradient in electron density profile, we can consider the sporadic E-layer where the electron density can reach much higher levels than the background ionosphere at E-layers.

Let us assume that the altitude profiles of the electron and ion temperatures are the same as Figure 3.2 (a) and (b). The electron density profile, however, changes according to Figure 3.11. We simulate the incoherent scatter measurements according to the procedure mentioned in Section 3.4.1. Similarly to the previous case, the data is integrated over 12000 pulse transmissions. The background noise added is also the same. To implement the hybrid technique, we use both Tikhonov and TV regularization techniques. Because the integration time and the background noise remain the same in data simulation, we use the optimal regularization parameter ($\lambda = 25$) obtained in Section 3.4.1. The estimation results are presented in Figure 3.12. The estimated electron density does not follow the sharp gradients of the original density profile. Instead, the estimated profile seems to be rather smooth.

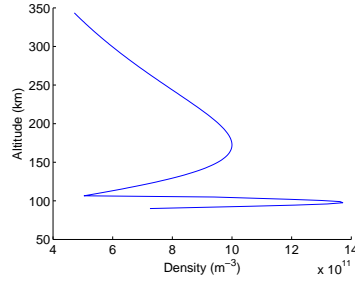


Figure 3.11 Electron density profile with sharp gradients.

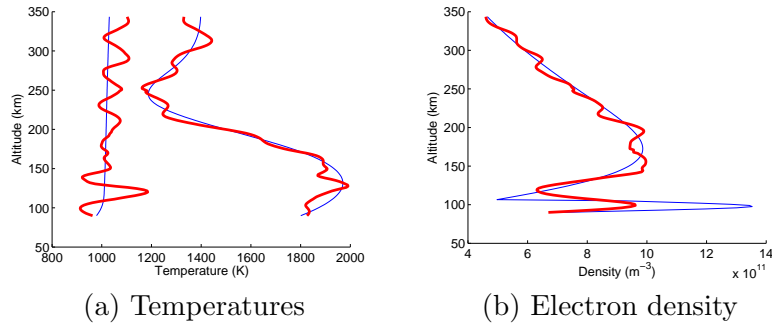


Figure 3.12 Tikhonov estimated ion and electron temperatures (a), and estimated electron density (b) in thick curves. Original profiles are shown in thin curves. The estimated electron density does not follow the sharp gradients of the original density profile.

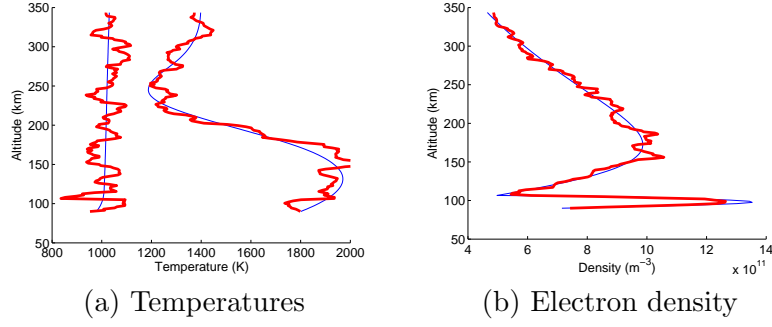


Figure 3.13 TV estimated ion and electron temperatures (a), and estimated electron density (b) in thick curves. Original profiles are shown in thin curves. The TV estimated electron density clearly follows the sharp gradients in the original density profile, while the errors in temperatures are nearly the same as those obtained via the Tikhonov technique.

Table 3.1 Summary of estimation errors for plasma parameters obtained with the Tikhonov and TV techniques.

	$T_e(K)$	T_i (K)	N_e (m^{-3})
Tikhonov	51	40	$7.5e^{10}$
TV	50	44	$2.8e^{10}$

The TV estimated profiles are shown in Figure 3.13. We use the same regularization parameter for the reconstruction. The estimated electron density with the TV technique clearly follows the sharp gradients in the original density profile, while the errors in temperatures are nearly the same as those obtained via the Tikhonov technique. Table 3.1 summarizes the estimation errors for both techniques, which shows a much higher error in the Tikhonov estimated electron density.

CHAPTER 4

INVESTIGATION OF CODING SCHEMES FOR ISR MEASUREMENTS

In the previous chapter, we presented an efficient hybrid technique for estimation of plasma parameters which attempts to remove the smearing of information from ISR measurements. However, based on our empirical results, we realize that the method is only able to partially remove the ambiguity and that the effective range resolution in long-pulse measurements is about half the distance covered by the transmitted pulse length. It is at this resolution that minimum parameter estimation errors occur. This coarse range resolution is the result of strong smearing of information which is induced in measurements by the long-pulse modulation.

The purpose of this chapter is to study and compare the performance of amplitude modulation with that of long-pulse modulation in terms of statistical accuracy and range resolution by means of numerical simulation. We show that the correlation between lag estimate errors is highly affected by the envelope of the modulated waveform for a given SNR, and that use of amplitude modulation yields significantly lower correlation. We investigate the effect of this correlation on statistical accuracy and the fundamental range resolution, which is supported by ISR measurements.

This chapter is organized as follows: In Section 6.2, we present analytical expressions for the correlation between error lag estimates, and show how this correlation is affected by the modulation waveform. We then investigate the effect of this correlation on ACF lag variance reduction when measurements are integrated in range or lag. Section 6.3 is dedicated to presenting analytical approaches to determine the fundamental resolution for long pulse in view of model order selection methods. Next, we present numerical estimates of the variance reduction and fundamental range resolution as a function of SNR for various AM waveforms with different on-off ratios.

4.1 Error covariance estimates

The received signal $v(t)$ is a sum of two signals, the radar echo $z(t)$ and the noise $n(t)$. The radar echo is a Gaussian stochastic process because it is the sum of a large number of independent elementary contributions within the scattering volume (central limit theorem [Papoulis, 1986]). Additionally, the echo is a wide sense stationary stochastic process (over a finite time when the ionosphere does not change significantly). The noise process is Gaussian, stationary and independent of the radar echo. Since both these processes have zero means and are Gaussian, they are specified by their correlation functions. Let us denote the correlation functions of the radar echo, the noise, and the received signal at time t and time lag $t' - t$ by $k(t, t')$, $k_n(t, t')$, and $K(t, t')$, respectively, as follows

$$k(t, t') = E[z(t)z(t')^*] \quad (4.1)$$

$$k_n(t, t') = E[n(t)n(t')^*] \quad (4.2)$$

$$K(t, t') = E[(z(t) + n(t))(z(t') + n(t'))^*] \quad (4.3)$$

$$= k(t, t') + k_n(t, t') \quad (4.4)$$

where $E[.]$ and $*$ denote the expected value and conjugation operations, respectively. In practice, however, we aim at obtaining estimates of the expected correlation function, $K(t, t')$, by calculating the cross products $(z(t) + n(t))(z(t') + n(t'))^*$ over repetitions of the experiment. In other words, we form the following sum, as ACF lag estimates, over different scans, when the same radar pulse is sent repeatedly

$$M(t, t') = \frac{1}{ND} \sum_{sc=1}^{ND} (z(t) + n(t))(z(t') + n(t'))^* \quad (4.5)$$

where sc denotes different scan counts and ND presents the total number of pulse transmissions. Since ND is finite and the ionosphere is continuously subject to change, these estimates of ACF lags deviate from the true values. These deviations can be defined as follows

$$\Delta M(t, t') = M(t, t') - E[M(t, t')] = M(t, t') - K(t, t') \quad (4.6)$$

In their work, *Lehtinen and Haggstrom* [1987] have shown that these deviations are not independent from one lag or one altitude to another, and that the correlation can be expressed in terms of the true expected functions, $k(t, t')$ and $k_n(t, t')$, or $K(t, t')$. Using the fourth-moment theorem for Gaussian processes, we can express the covariance between the real (Re) and imaginary (Im) parts of the errors as [*Lehtinen and Huuskonen*, 1996]

$$E[\text{Re}\Delta M(t, t')\text{Re}\Delta M(u, u')] = \frac{1}{2ND}\text{Re}(K(t, u)K(u', t') + K(t, u')K(u, t')) \quad (4.7)$$

$$E[\text{Im}\Delta M(t, t')\text{Im}\Delta M(u, u')] = \frac{1}{2ND}\text{Re}(K(t, u)K(u', t') - K(t, u')K(u, t')) \quad (4.8)$$

$$E[\text{Re}\Delta M(t, t')\text{Im}\Delta M(u, u')] = \frac{1}{2ND}\text{Im}(-K(t, u)K(u', t') + K(t, u')K(u, t')) \quad (4.9)$$

$$E[\text{Im}\Delta M(t, t')\text{Re}\Delta M(u, u')] = \frac{1}{2ND}\text{Im}(K(t, u)K(u', t') - K(t, u')K(u, t')) \quad (4.10)$$

These expressions imply that the covariance between the estimates $M(t, t')$ and $M(u, u')$ does not depend explicitly on the corresponding expected plasma ACFs, $K(t, t')$ and $K(u, u')$, but rather on the expectation of the ACF evaluated at the other four possible permutations. Therefore, evaluation of the full error covariance matrix becomes more computationally expensive, especially in high SNR scenarios where the matrix becomes far from diagonal. For example, the errors between ACF from adjacent heights or adjacent lags are always correlated because of zeroth lag terms that will appear in error expressions (4.7) to (4.10).

In order to visualize the effect of the transmitted waveform on correlation between errors, in Figure 4.1 we illustrate the error correlation coefficients of lags from one altitude for a long pulse and several amplitude modulated pulses with different duty cycles (ranging from 1 to 0.75 to 0.64 to 0.57 to 0.42 from left to right) shown in Figure 4.2. Note also that lag resolution is set to $10 \mu\text{s}$ and a hypothetical ionosphere (oxygen only with no drift velocity) and constant parameter profiles ($T_i = T_e = 1000 \text{ K}$) are assumed. From the plots in Figure 4.1, it is clear that as the duty cycle of a code decreases, the correlation length reduces to fewer lag neighbors. Note that the modulations used in this simulation are chosen such that they provide information about all ACF lags with $10 \mu\text{s}$ resolution up to $250 \mu\text{s}$, and AM4 is the five-pulse modulation (see [*Farley*, 1972]) with $20 \mu\text{s}$ lag resolution.

When the data is integrated over a few lags (ranges), the reduction in ACF lag variance becomes more pronounced for codes with lower duty cycles. Figure 4.3 shows this error variance reduction

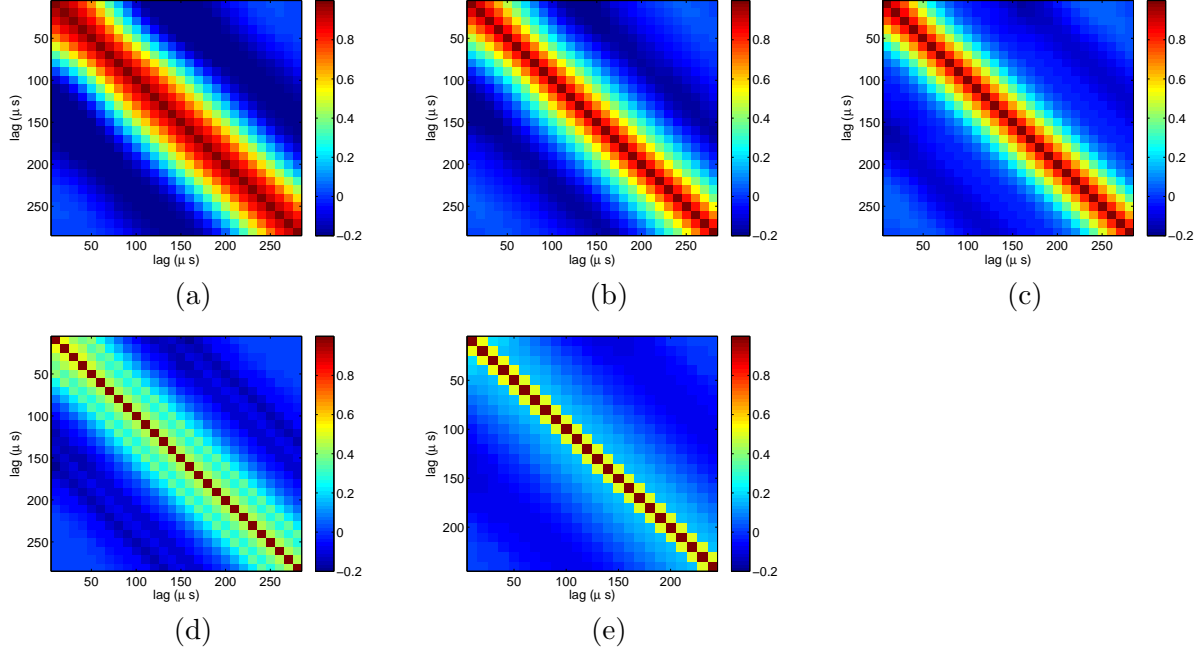


Figure 4.1 Correlation coefficients between lag estimate errors for amplitude modulated pulses in Figure 4.2.

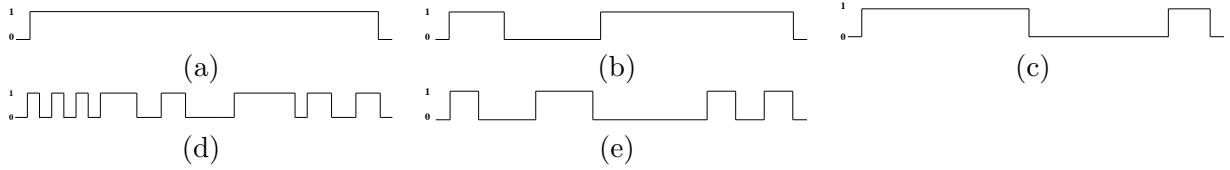


Figure 4.2 Modulations used in the simulations. Duty cycles are 1, 0.75, 0.64, 0.57, and 0.42 for pulses in (a)-(e), respectively.

when data is integrated over three consecutive lags as a function of both SNR and spectrum width (ion temperature). In long-pulse measurements where the errors are highly correlated, the integration of signal over three ranges results in variance reduction (improvement in estimation accuracy) by a factor of 1.1 instead of 3 in the case of fully independent observations. The improvement in estimation accuracy becomes more evident as the on-off ratio of the modulated pulse is decreased.

Figure 4.4 shows the reduction in ACF lag variance as a function of both SNR and spectrum width when measurements with different modulation techniques are integrated over three consecutive ranges (separated by $10 \mu\text{s}$). At $\text{SNR} = 1$, compared to the long pulse, AM3 and AM4 show 52% and 63% improvement in statistical accuracy, respectively. This essentially means that compared to the long pulse, the integration time required to achieve a certain level of estimation error for a fixed range resolution in high SNR ($\text{SNR} > 0.4$) is 52% and 63% shorter with AM3 and AM4,

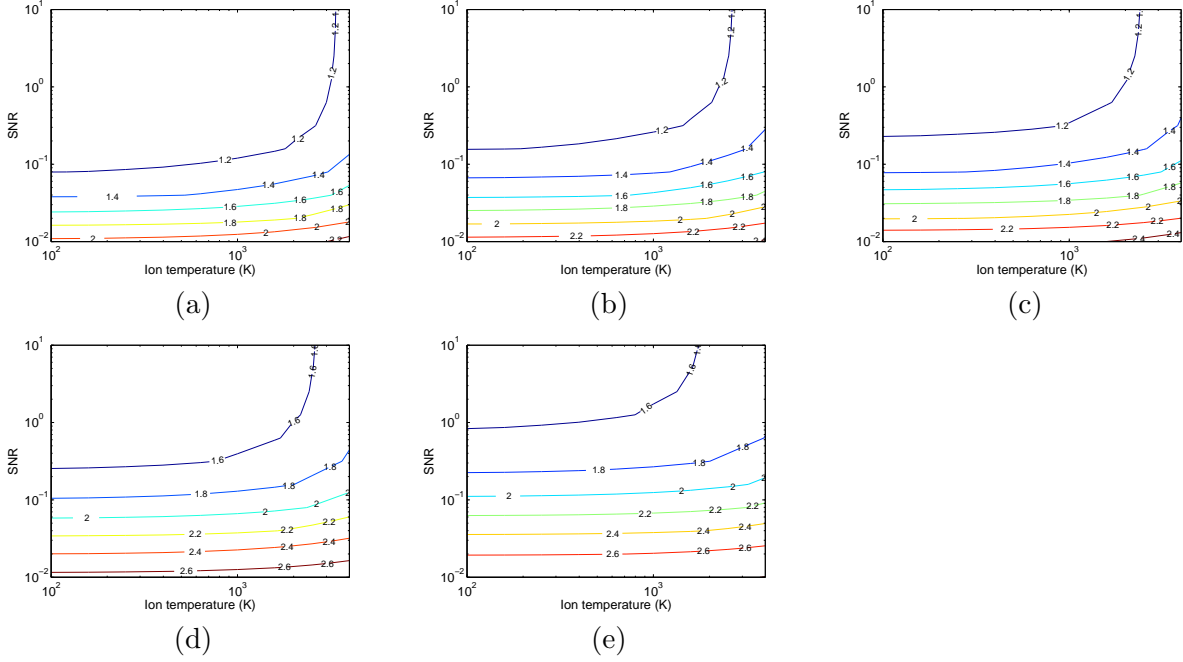


Figure 4.3 Improvement in statistical accuracy due to reduced variance in over-sampling for five pulses in Figure 4.2 as a function of spectrum width (ion temperature) and SNR (data are integrated for three consecutive lags). The improvement in accuracy which can be directly translated to improvement in integration time can be up to 45% with respect to the long pulse.

respectively. Equivalently, for a fixed integration time, same estimation error levels will be obtained over finer range resolution grids. This issue is addressed from a model order selection standpoint in the next section.

It should be mentioned that the duty cycle of amplitude-modulated codes affects sensitivity to the noise background. An example of relative errors in estimated zeroth lag for pulses with different duty cycles as a function of SNR is presented in Figure 4.5. The same level of relative error is reached at higher SNRs as the on-off ratio of the code decreases, confirming the higher sensitivity of measurements to background and receiver noise for pulses with smaller on-off ratios. In addition, in lower duty cycles, usually one or more lags become missing, and this lack of certain ACF lags can introduce large errors in estimation. Therefore, in this work, we consider 0.41 as the lowest duty cycle for the AM pulses .

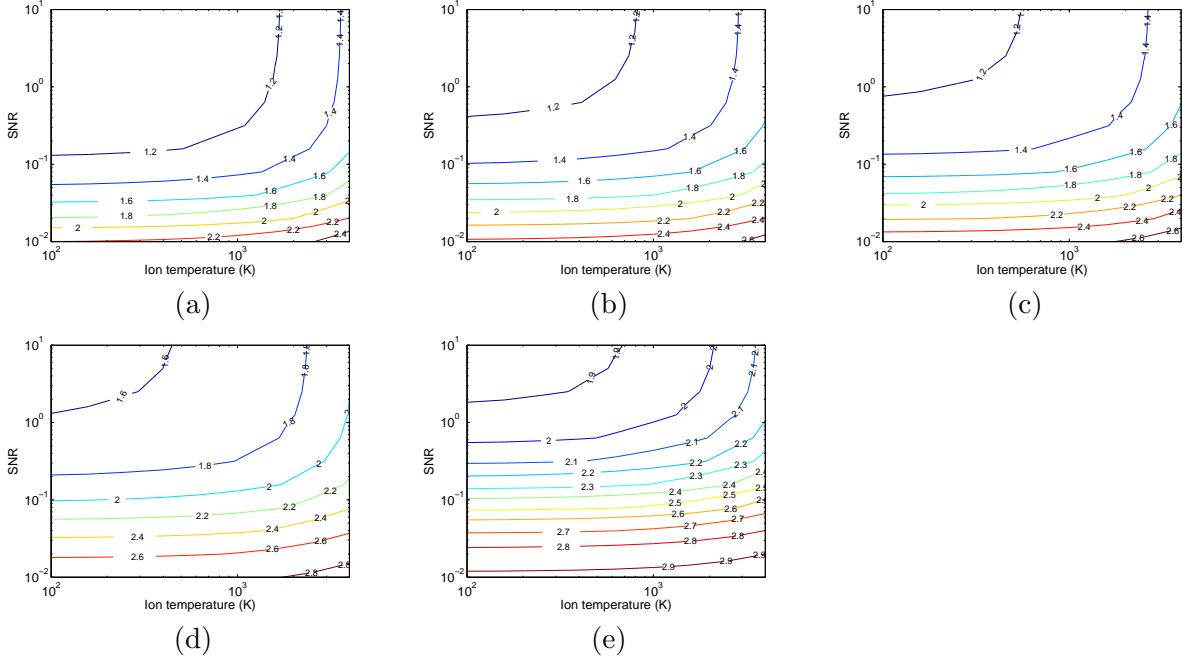


Figure 4.4 Improvement in statistical accuracy due to reduced variance when the data are integrated over three consecutive ranges for five pulses in Figure 4.2 as a function of spectrum width and SNR. The improvement in accuracy which can be directly to improvement in integration time can be up to 65% with respect to the long pulse.

4.2 Fundamental resolution in ISR measurements

4.2.1 Model order selection framework

In this section we develop a framework, based on model order selection, for quantifying resolution in ISR measurements. This problem essentially can be expressed as finding the optimal grid density over which the physical parameters are most accurately estimated from the measurements. In the hybrid technique, models of plasma ACF lag profiles are viewed as discretization grids with different densities.

In the hybrid technique, the physical parameters are estimated by first computing plasma ACF lag profiles (denoted by \mathbf{m}_τ for $\tau = 0, \Delta\tau, \dots$, where τ and $\Delta\tau$ represent time lag and time lag increment, respectively) from the received signal ACF (denoted by \mathbf{y}_τ). To find the optimal resolution, we consider the discretization of the plasma ACF lag profile at lag τ , \mathbf{m}_τ , as

$$\mathbf{m}_\tau = \sum_{i=1}^v \theta_\tau^i \phi^i \quad (4.11)$$

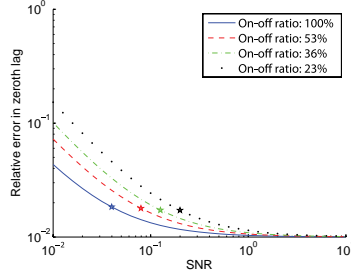


Figure 4.5 Relative errors in the zeroth lag ACF as a function of SNR for pulses with various duty cycles.

which is equivalent to parametrization of \mathbf{m}_τ with v known basis functions $\{\phi^i\}_{i=1}^v$ and the corresponding unknown coefficients $\{\theta_\tau^i\}_{i=1}^v$. In this work, we consider $\{\phi^i\}$ as the set of unit rectangular boxes which constitutes an ortho-normal basis for the plasma ACF lag profiles. Different models can then be considered as a combination of unit rectangular boxes with different widths. Here, the larger the widths of these basis functions, the lower the number of required functions to specify each ACF lag profile and hence the coarser the resolution. The representation of functions in terms of orthogonal systems of functions is a standard mathematical technique which has been employed by statisticians for estimating probability distributions (e.g. see [Kendell and Stuart, 1979; Sharif and Kamalabadi, 2005]).

Note that the discretization grid density for the measurement ACF is fixed and determined by the time lag increment. According to the development in Section 3.2, when the grid density of \mathbf{m}_τ is also determined by $\Delta r = \frac{\Delta \tau \times c}{2}$ (where c is the speed of light), the relationship between the two sets of ACFs is expressed in the matrix framework as

$$\mathbf{y}_\tau = \mathbf{P}_\tau \boldsymbol{\theta}_\tau + \epsilon_\tau \quad (4.12)$$

where \mathbf{P}_τ is the convolution matrix constructed based on the ambiguity function at a given time lag τ and ϵ represents the measurement errors. In (4.12), \mathbf{y}_τ and $\boldsymbol{\theta}_\tau$ are $n \times 1$ vectors, where n is the number of altitudes of interest (with spacing equal to Δr) and \mathbf{P}_τ is an $n \times n$ matrix. We can further combine all such equations for all time lags and form a unified matrix equation as follows

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \epsilon \quad (4.13)$$

where the “projection matrix (forward model)” \mathbf{A} is a block diagonal matrix with each convolution matrix as its diagonal blocks, given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_\tau & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ & & & \mathbf{P}_{\frac{T}{\Delta\tau}} \end{bmatrix} \quad (4.14)$$

with T the pulse length. The measurement and unknown vectors can be represented by $\mathbf{y} = [\mathbf{y}_0; \mathbf{y}_{\Delta\tau}; \cdots]$ and $\boldsymbol{\theta} = [\boldsymbol{\theta}_0; \boldsymbol{\theta}_{\Delta\tau}; \cdots]$, respectively. In this case, the data and unknowns are $(n \times \frac{T}{\Delta\tau}) \times 1$ vectors and \mathbf{A} is a $(n \times \frac{T}{\Delta\tau}) \times (n \times \frac{T}{\Delta\tau})$ matrix.

For other models with different grid densities, the relationship between the measurements and unknowns is still representable in a similar fashion. The only difference, however, is in projection matrices (forward models) and representation of the unknowns in terms of basis functions. As the resolution becomes coarser, the number of the unknowns decreases while the number of data points remains the same. For example, if the resolution becomes k times coarser, i.e. the width of the basis function $\phi_{(k)}$ becomes k times larger, the unknown vector $\boldsymbol{\theta}_{(k)}$ becomes $(n \times \frac{T}{k\Delta\tau}) \times 1$ (as opposed to $(n \times \frac{T}{\Delta\tau}) \times 1$). The projection matrix $\mathbf{A}_{(k)}$ is then obtained by averaging each k columns of \mathbf{A} to yield a $(n \times \frac{T}{\Delta\tau}) \times (n \times \frac{T}{k\Delta\tau})$ matrix. Note that the subscript $_{(k)}$ denotes the resolution which is coarser by a factor of k .

With the assumption of the Gaussian errors, the probability distribution of the model k can be described as

$$f_{(k)}(\mathbf{y}|\boldsymbol{\theta}_{(k)}) = \frac{1}{(2\pi|\Sigma_\epsilon|)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\|\mathbf{y} - \mathbf{A}_{(k)}\boldsymbol{\theta}_{(k)}\|_{\Sigma_\epsilon^{-1}}^2\right) \quad (4.15)$$

where Σ_ϵ represents the error covariance matrix of measurements.

4.2.2 Model order selection method

Having identified the set of models for the discretization of ACF lag profiles, we now need to select a good model that approximates the true structure as accurately as possible through the use of available data. To do so, one has to consider a model evaluation criterion and choose a model that

optimizes that criterion. The model evaluation criterion that we consider here is an “information criterion” which has been widely used in the model order selection literature (see e.g. [Konishi and Kitagawa, 2008]). The information criterion is defined in terms of Kullback-Leibler (K-L) information (difference between the predictive distribution defined by the model $f_{(k)}$ and the true distribution g , $I(g, f_{(k)}) = \int g(\mathbf{y}) \log \frac{g(\mathbf{y})}{f_{(k)}(\mathbf{y})} d\mathbf{y}$) as an indicative of the “information” lost when $f_{(k)}$ is used to approximate g [Cover and Thomas, 1991; Akaike, 1973; Konishi and Kitagawa, 2008]. The K-L distance can be simplified as follows

$$I(g, f_{(k)}) = \int g(\mathbf{y}) \log(g(\mathbf{y})) d\mathbf{y} - \int g(\mathbf{y}) \log(f_{(k)}(\mathbf{y}|\boldsymbol{\theta}_{(k)})) d\mathbf{y} \quad (4.16)$$

$$= E_g[\log[g(\mathbf{y})]] - E_g[\log[f_{(k)}(\mathbf{y}|\boldsymbol{\theta}_{(k)})]] \quad (4.17)$$

where $E_g[\cdot]$ denotes the expectation with respect to the true distribution g . The first expectation, $E_g[\log[g(\mathbf{y})]]$, is a constant that depends only on the unknown true distribution, and is clearly not known. With treating this unknown as a constant (C), only a measure of *relative* K-L distance is possible up to the constant C because

$$I(g, f_{(k)}) = C - E_g[\log[f_{(k)}(\mathbf{y}|\boldsymbol{\theta}_{(k)})]] \quad (4.18)$$

In model selection based on the information criterion, in the first step we obtain unknown parameters for each candidate parametric model using the maximum likelihood estimation or other estimation methods such as robust or penalized likelihood estimation methods. We next construct a statistical model $f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}}_{(k)})$ by replacing the unknown parameter $\boldsymbol{\theta}_{(k)}$ contained in the probability distribution of the parametric model by its estimator $\hat{\boldsymbol{\theta}}_{(k)}$. The goodness (or badness) of each model then is determined by $-E_g[\log[f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}}_{(k)})]]$. Although this quantity is not computable (since the true model g is not known), it can be estimated by the log of likelihood of each model, presented as [Burnham and Anderson, 1998]

$$\log(\mathcal{L}(\hat{\boldsymbol{\theta}}(\mathbf{y})|\mathbf{y})) \equiv \log(f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}}(\mathbf{y}))) \quad (4.19)$$

which is directly available to us. However, it is shown that in the maximum likelihood framework the estimator in (4.19) is biased upward where the bias correction term can be represented by

$\text{Tr}(\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1})$ [Akaike, 1973]. In this bias term, $\text{Tr}(\cdot)$ represents the trace operation, and $\mathbf{J}(\hat{\boldsymbol{\theta}})$ and $\mathbf{I}(\hat{\boldsymbol{\theta}})$ involve the first- and second-order derivatives of the likelihood function with respect to the parameter of interest $\boldsymbol{\theta}$ evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, as follows

$$\mathbf{I}(\hat{\boldsymbol{\theta}}) = \left\{ -\frac{\partial^2 \log(f_{(k)}(\mathbf{y}|\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j} \right\}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (4.20)$$

$$\mathbf{J}(\hat{\boldsymbol{\theta}}) = \left\{ \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log(f_{(k)}(\mathbf{y}|\boldsymbol{\theta})) \right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log(f_{(k)}(\mathbf{y}|\boldsymbol{\theta})) \right]' \right\}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (4.21)$$

Akaike [1973] also shows that under certain conditions (if the true model belongs to the set of candidate models), this bias is approximately equal to K , the number of to-be-estimated parameters in the approximating models. Thus, an approximately unbiased estimator of the relative K-L distance becomes

$$-\log(\mathcal{L}(\hat{\boldsymbol{\theta}}(\mathbf{y})|\mathbf{y})) + K = \text{estimated relative expected K-L distance} \quad (4.22)$$

$$= -\log(f_{(k)}(\mathbf{y}|\boldsymbol{\theta})) + K \quad (4.23)$$

In the penalized likelihood (regularization) framework, such as the estimation of plasma ACF lag profiles from ISR data in the hybrid technique, the bias term will differ from the number of to-be-estimated parameters. To develop this term, one needs to consider the modified log-likelihood function as the Tikhonov regularization [Nikoukar *et al.*, 2008]

$$l_{\lambda}(\hat{\boldsymbol{\theta}}) = \log f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}}) - \lambda^2 \hat{\boldsymbol{\theta}}' \mathbf{L}' \mathbf{L} \hat{\boldsymbol{\theta}} \quad (4.24)$$

where λ and \mathbf{L} are regularization parameter and functional, respectively. With the current setting, the estimation solution, $\hat{\boldsymbol{\theta}}$, can be obtained by solving the following equations

$$\begin{aligned} \psi_p(\mathbf{y}, \hat{\boldsymbol{\theta}}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}}) - \lambda^2 \boldsymbol{\theta}' \mathbf{L}' \mathbf{L} \boldsymbol{\theta} \right\} \\ &= 0 \end{aligned} \quad (4.25)$$

And the generalized information criterion (GIC) for model estimated by regularization is obtained

by [Konishi and Kitagawa, 2008]

$$GIC = -2\log(f_{(k)}(\mathbf{y}|\hat{\boldsymbol{\theta}})) + 2Tr\{\mathbf{R}(\psi_p)^{-1}\mathbf{Q}(\psi_p)\} \quad (4.26)$$

where $\mathbf{R}(\psi_p)$ and $\mathbf{Q}(\psi_p)$ are $(n \times \frac{T}{k\Delta\tau}) \times (n \times \frac{T}{k\Delta\tau})$ matrices, respectively, given by

$$\mathbf{R}(\psi_p) = -\frac{\partial\psi_p(\mathbf{y}, \boldsymbol{\theta})^T}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (4.27)$$

$$\mathbf{Q}(\psi_p) = \psi_p(\mathbf{y}, \boldsymbol{\theta})\frac{\partial\log f(\mathbf{y}, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (4.28)$$

For the problem at hand, $\mathbf{R}(\psi_p)$ and $\mathbf{Q}(\psi_p)$ are given by

$$\mathbf{R}(\psi_p) = (\mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A} + \lambda^2\mathbf{L}^T\mathbf{L}) \quad (4.29)$$

$$\mathbf{Q}(\psi_p) = \mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A} \quad (4.30)$$

The bias term on the right-hand side of (4.26) is expressed as

$$\text{Tr}\{\mathbf{R}(\psi_p)^{-1}\mathbf{Q}(\psi_p)\} = \text{Tr}\{(\mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A} + \lambda^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A}\} \quad (4.31)$$

$$= \text{Tr}\left(\Sigma_\epsilon^{-1/2}\mathbf{A}(\mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A} + \lambda^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{A}^T\Sigma_\epsilon^{-1/2}\right) \quad (4.32)$$

$$= \text{Tr}(\mathbf{M}_\lambda) \quad (4.33)$$

where $\mathbf{M}_\lambda = \Sigma_\epsilon^{-1/2}\mathbf{A}(\mathbf{A}^T\Sigma_\epsilon^{-1}\mathbf{A} + \lambda^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{A}^T\Sigma_\epsilon^{-1/2}$. The term \mathbf{M}_λ relates the predictive and actual data ($\hat{\mathbf{y}}$ and \mathbf{y} , respectively), as $\hat{\mathbf{y}} = \mathbf{M}_\lambda\mathbf{y}$, and is called the *influence* or *smoothing* matrix. The trace of this matrix is always less than the number of parameters and indicates the *effective number of parameters* or *degree of freedom* (*df*). This reduction in the number of effective parameters is due to regularization which brings some dependencies into neighboring parameters.

4.3 Numerical estimates of range resolution

Here, we present the results of numerical resolution analysis for the five modulations shown in Figure 4.2 for a fixed integration time (100 μs). Here, we set the length of transmitted pulse and the time lag increment to 300 μs and 10 μs , respectively, with the original physical parameter

profiles shown in Figure 3.2. We also consider three different SNR regimes (high, $\text{SNR} > 0.4$; medium, $0.08 < \text{SNR} < 0.2$; and low, $\text{SNR} < 0.07$) for which the altitude profiles of SNR are shown in Figure 4.6 (the details on data simulation have been explained in Chapter 3).

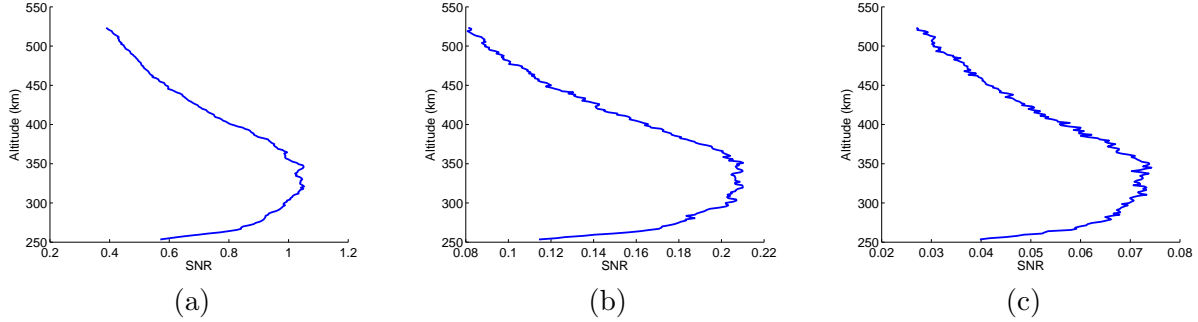


Figure 4.6 Signal-to-noise ratio profiles used for ISR numerical experiments.

We first start with the estimation of range resolution for long-pulse measurements in the high SNR regime (see Figure 4.6(a)). In order to determine the range resolution, we consider the models with basis functions with widths 10, 20, 30, 50, 60, 100, 150, and 300 μs , corresponding to 1.5, 3, 4.5, 7.5, 9, 15, 22.5, and 45 km, respectively. (These are noted by models $A_{(1)}$, $A_{(2)}$, $A_{(3)}$, $A_{(5)}$, $A_{(6)}$, $A_{(10)}$, $A_{(15)}$, and $A_{(16)}$, respectively. Model $A_{(1)}$ has the finest resolution, and $A_{(k)}$ is the model with resolution k times coarser.)

For each model, the regularized solution is obtained based on (3.11) for various regularization parameters, and the expression of GIC is computed from (4.26) for each model and each regularization parameter. The bias term that we use in the analysis is the effective number of parameters presented by (4.31). Figure 4.7 shows values of GIC for the model $A_{(1)}$ for which the width of the basis function is equal to the lag resolution (1.5 km). As expected, GIC first decreases with increasing regularization parameter and then increases. This indicates that GIC can be also used for the selection of the optimum regularization parameter for a fixed model. The smallest parameter which results in minimized GIC is depicted by an “o” in Figure 4.7 (a).

Once the optimum regularization is determined, we can obtain the true resolution based on $\text{Tr}(\mathbf{M}_\lambda)$, the trace of the corresponding influence matrix. As explained previously the trace of the influence matrix for a specific λ is a representative of the number of free parameters in the plasma

ACF lag estimation problem. The optimal resolution is obtained as follows

$$Resolution = \frac{n \times \frac{T}{k\Delta\tau}}{\text{Tr}(\mathbf{M}_\lambda)} \times \text{width}(\phi_{(k)}) \quad (4.34)$$

where $n \times \frac{T}{k\Delta\tau}$ is the total number of unknowns in the lag estimation problem when model $A_{(k)}$ is used. In a hypothetical case where the measurements are noise-free, no regularization is necessary and the number of free parameters is equal to the number of unknowns for each model. In this case, the resolution is equal to the width of the model basis function, as expected. In realistic cases, however, df is considerably smaller than the number of unknowns, and, hence, the resolution differs significantly from the width of the basis function for long-pulse measurements. Figure 4.7(b) shows the GIC values as a function of the resolution for the long-pulse measurements. The optimal resolution shown by an “o” is approximately 23 km, which is nearly half of the distance covered by the length of the transmitted pulse. This result is in agreement with that presented in [Lehtinen and Huuskonen, 1996] where the optimal grid density was obtained via a full-profile analysis technique.

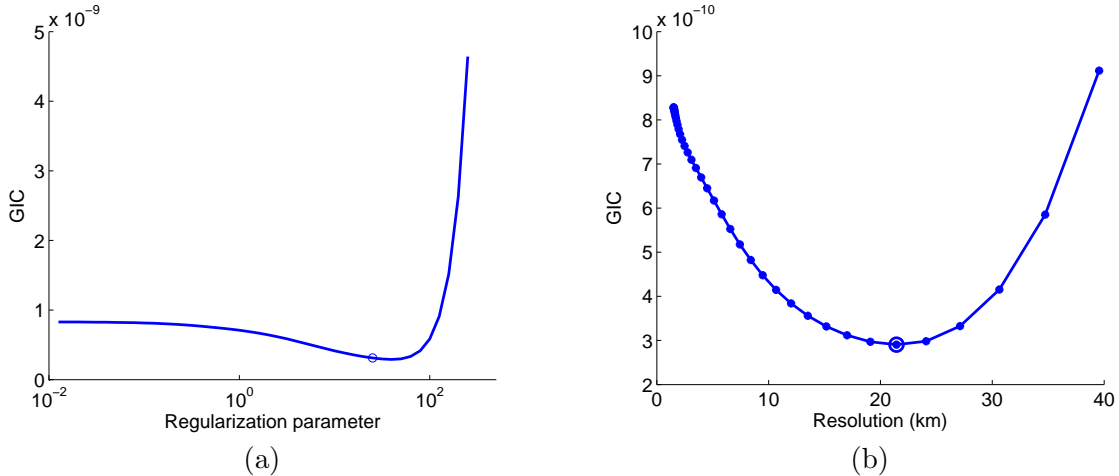


Figure 4.7 GIC values for long-pulse measurements for the data with SNR shown in Figure 4.6(a) as a function of regularization parameter (a) and resolution (b). The optimal regularization parameter and optimal resolution are shown by an “o” in (a) and (b), respectively.

Figure 4.8(a) shows the variation of GIC for models $A_{(1)}$, $A_{(2)}$, $A_{(3)}$, $A_{(5)}$, $A_{(6)}$, $A_{(10)}$, and $A_{(15)}$ as a function of range resolution. As in the previous case, for all these models, GIC first decreases with an increase in regularization parameter and then increases, with minimum GIC occurring at resolution close to 23 km. We note here that the minimum GIC decreases as we consider models with lower grid densities. The minimum of these values occurs for model $A_{(15)}$ when regularization

parameter is set to 0, and the resolution is 22.5 km. This result indicates that although the number of unknown parameters differs in different models, the effective number of parameters, and hence the resolution, remains nearly constant for all models. The reason is that the true resolution is implicit in the expression of GIC through the bias term. The GIC for $A_{(30)}$ is presented in Figure 4.8(b). As seen in this figure, GIC for all regularization parameters (resolution) is significantly higher than that for other models. The smallest resolution offered by this model is equal to the length of the transmitted pulse, which is much coarser than the optimal resolution. This implies that this model is not appropriate as a deconvolution model. This explains the poor performance of the height-by-height analysis method, where the physical parameters are estimated with spacing equal to the length of the transmitted pulse (see Section 3.4).

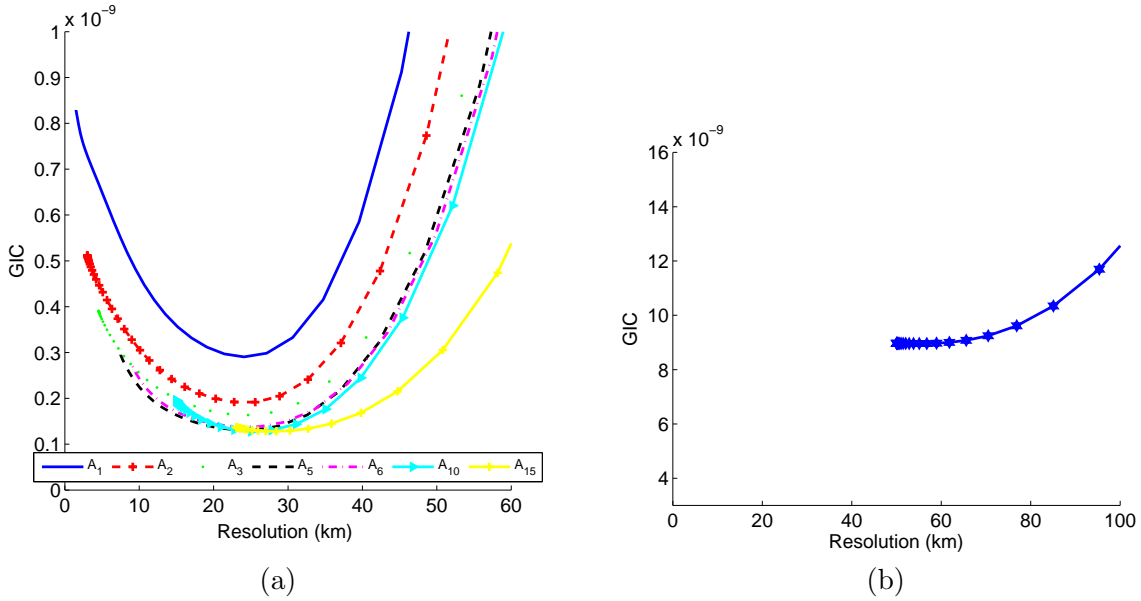


Figure 4.8 GIC values for long-pulse measurements as a function of range resolution for models $A_{(1)}$, $A_{(2)}$, $A_{(3)}$, $A_{(5)}$, $A_{(6)}$, $A_{(10)}$, and $A_{(15)}$ in (a) and $A_{(30)}$ in (b). All models except $A_{(30)}$ offer an optimal resolution at nearly 23 km. The resolution achievable by $A_{(30)}$ is too coarse even with no regularization applied.

We now focus on estimation of range resolution for other amplitude modulations and SNR regimes shown in Figures 4.2 and 4.6, respectively. For analysis we consider GIC values for one model only ($A_{(1)}$ with the highest density). Because we use regularization, the effective number of parameters for the optimum regularization parameters for the highest density model coincides with the number of independent variables for one of the models with lower densities.

Table 4.1 Summary of resolutions achievable with different modulation schemes and different SNR levels. Resolutions are expressed in terms of the transmitted waveform length.

	$\text{SNR} > 0.4$	$0.08 < \text{SNR} < 0.2$	$\text{SNR} < 0.07$
Long pulse	0.5	0.5	0.5
AM1	0.38	0.38	0.38
AM2	0.29	0.35	0.35
AM3	0.2	0.22	0.3
AM4	0.17	0.22	0.3

Plots in Figure 4.9(a)-(d) show GIC values as a function of range resolution for ISR measurements with AM1-AM4, respectively, for the high SNR scenario ($\text{SNR} > 0.4$ in Figure 4.6(a)). By comparison of these plots with those in Figure 4.7(b), we realize that as the on-off ratio of the modulated waveform decreases (from 1 to 0.41), finer range resolutions become achievable from ISR measurements (from 22.5 km to 8.3 km).

We repeat the same analysis for medium and low SNR scenarios (Figure 4.6) with results shown in Figures 4.10 and 4.11 and summarized in Table 4.1. It can be seen that resolutions for the long-pulse modulation and AM1 (with on-off ratio = 0.75) remain constant for the three SNR regimes at 22.5 km and 17.1 km, respectively. For AM2, the resolution first increases from 13 km to 15.7 km with SNR dropping from high to medium and then remains the same. For AM3, the resolution increases from 9 to 10 to 14.1 km for the three SNR regimes. In the case of measurements with AM4, range resolution increases from 8.3 km to 14 km with lowering SNR. The reason for variation in the optimal range resolution for modulations with smaller duty cycles is that when SNR is low, the relative errors in estimated ACF lags become higher, which requires higher regularization parameters to stabilize the deconvolution solutions, and hence results in coarser optimal range resolution.

Note that the range resolution for AM3 and AM4 is very similar in medium and low SNR regimes. This shows that lowering the on-off ratio of the modulated pulse not only has no significant effect on improving range resolution, but also makes the measurements more sensitive to the background noise level. This is an important result which should be considered in optimal amplitude modulated waveform design.

Once the optimal resolution is determined, we construct a forward model whose basis function

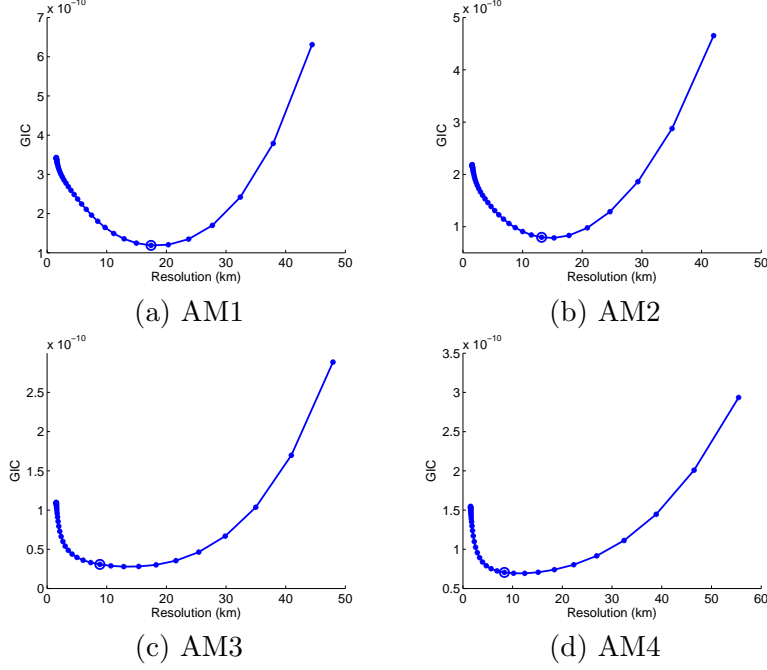


Figure 4.9 GIC as a function of range resolution for the SNR profile shown in Figure 4.6(a) for AM1-AM4.

width is equal to the resolution. Deconvolution, then, is obtained using least-squares techniques. No regularization is necessary in this case, as the effect of regularization parameters is already taken into account in the choice of the optimal model. To visualize how this mechanism works, we present estimated physical parameters using long-pulse and AM3 measurements for the high and low SNR cases. According to the results of Table 4.1, $A_{(15)}$ is the optimal model for long-pulse measurements in both cases, whereas $A_{(6)}$ and $A_{(10)}$ are optimal models for AM3 measurements in high and low SNR scenarios, respectively. Figures 4.12 and 4.13 show the estimation results. Based on these plots and Tables 4.2 and 4.3, we realize that although the range resolutions differ significantly especially for the high SNR level, estimated parameter errors remain nearly equal. Note that this result is the direct consequence of improved statistical accuracy for integrated data in lag and range directions which is more pronounced for amplitude modulation with lower on-off ratios. This is advantageous in situations where we need high resolution and a low integration time in a high or medium SNR scenarios and high-resolution techniques such as multipulse or alternating codes cannot be utilized.

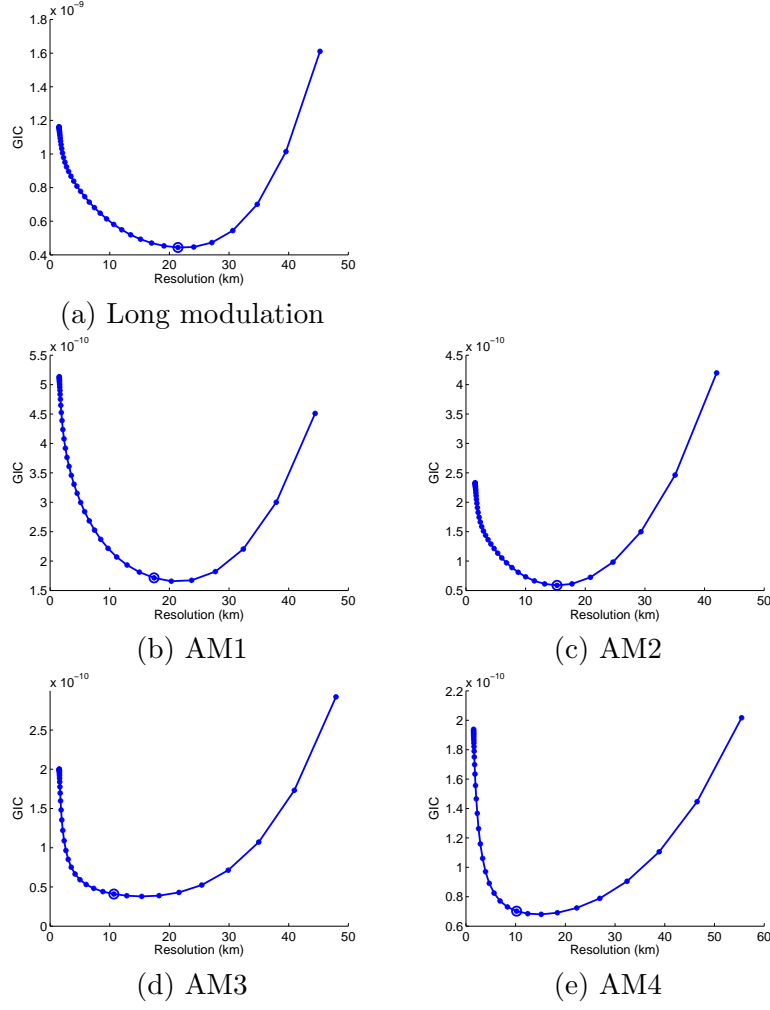


Figure 4.10 GIC as a function of range resolution for the SNR profile shown in Figure 4.6(b) for long pulse and AM1-AM4.

It is worth mentioning here that considering a low-resolution model for deconvolution can reduce computations significantly by reducing the number of unknown parameters while producing accurate results. This reduction in computation can be extremely useful in real-time data analysis.

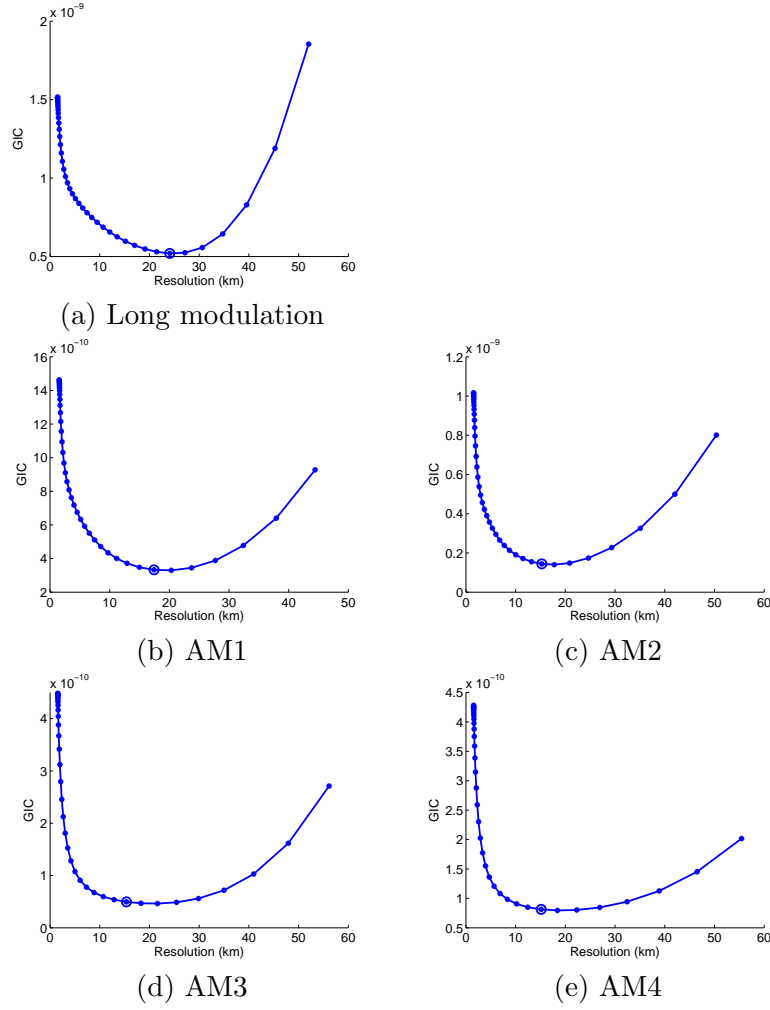


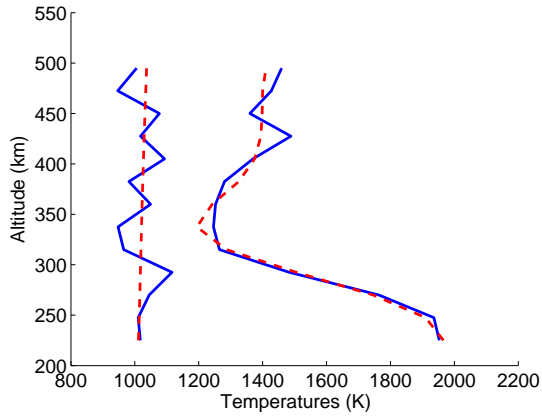
Figure 4.11 GIC as a function of range resolution for the SNR profile shown in Figure 4.6(c) for long pulse and AM1-AM4.

Table 4.2 Estimated errors in physical parameters for a high SNR situation for 50 runs of Monte Carlo simulations.

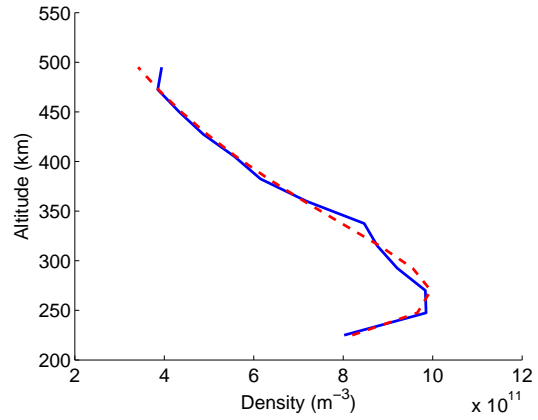
	Electron temperature (K)	Ion temperature (K)	Electron density (m^{-3})
Long pulse	29.1	38.2	2.0e10
AM3	30.2	40.1	1.9e10

Table 4.3 Estimated errors in physical parameters for a low SNR situation for 50 runs of Monte Carlo simulations.

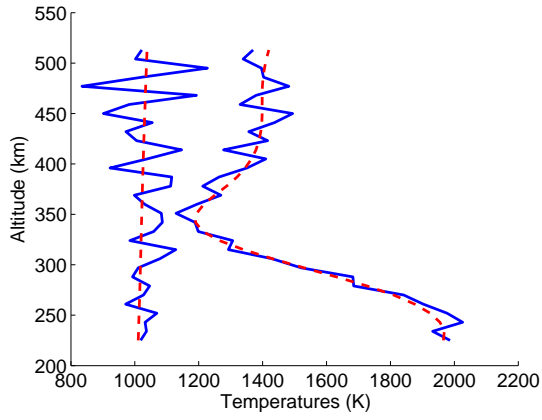
	Electron temperature (K)	Ion temperature (K)	Electron density (m^{-3})
Long pulse	43.2	48.6	1.9e10
AM3	44.7	47.1	1.9e10



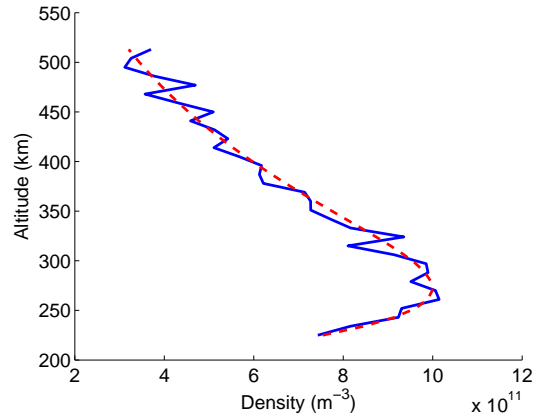
(a) Temperatures



(b) Electron density

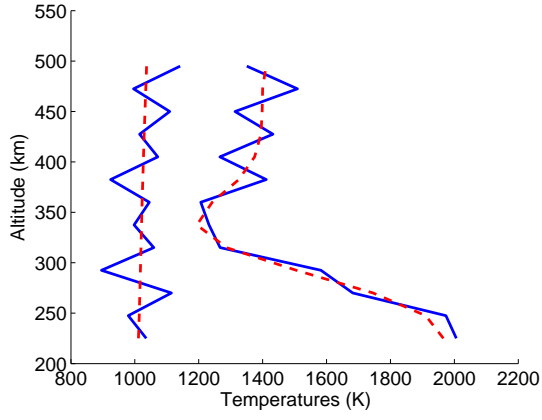


(a) Temperatures

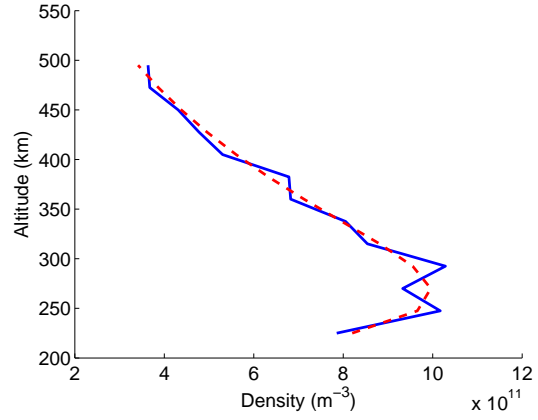


(b) Electron density

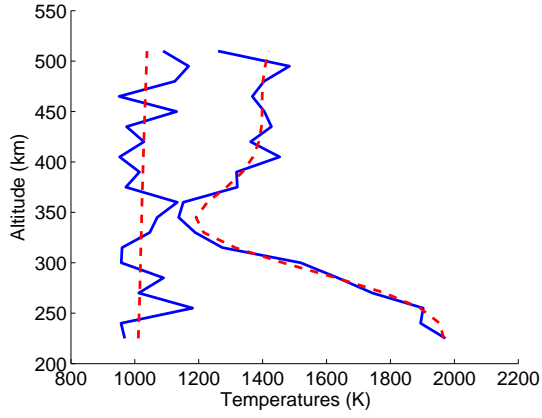
Figure 4.12 Estimated physical parameters (solid lines) for long-pulse (top) and AM3 (bottom) measurements in a high SNR regime. The optimal resolution for AM3 is improved by a factor of 2.5, while the estimation errors are nearly the same. Original parameters are denoted by dashed lines.



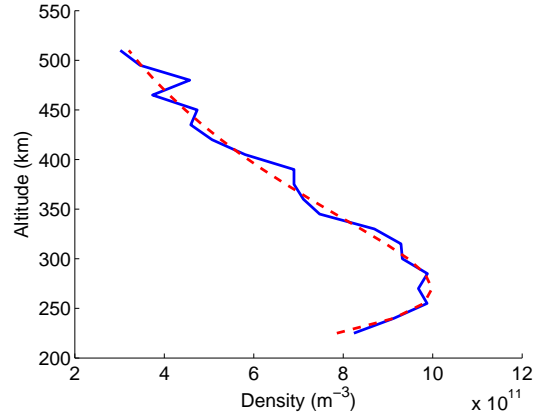
(a) Temperatures



(b) Electron density



(a) Temperatures



(b) Electron density

Figure 4.13 Estimated physical parameters (solid lines) for long-pulse (top) and AM3 (bottom) measurements in a low SNR regime. The optimal resolution for AM3 is improved by a factor of 1.5, while the estimation errors are nearly the same. Original parameters are denoted by dashed lines.

CHAPTER 5

EXPERIMENTAL INVESTIGATION OF CODING SCHEMES WITH ARECIBO ISR

The goal of this chapter is to describe several experiments conducted in July 2006, August 2005, and April 2004 using the incoherent scatter radar at Arecibo Observatory, and present the parameter estimation results obtained from them. These experiments include F-region ionospheric correlation measurements using a modified version of the multi-radar auto-correlation function (MRACF) mode and are designed to investigate the performance of amplitude modulated coding schemes with different duty cycles in terms of estimation accuracy and spatial resolution. The experimental results from the modified mode confirm that, compared with an unmodulated long-pulse, improved range resolution with nearly the same statistical accuracy can be obtained when an amplitude modulation is utilized in transmission.

This chapter is organized as follows. First, we describe the specific mode of MRACF for F-region measurements and the rationale behind it. We then discuss our proposed modifications. Next, we present ionospheric parameter estimation results for various coding schemes for each of the previously mentioned ISR experiments.

5.1 F-region measurements at Arecibo

The Arecibo radar is located at 18° N and 67° W, near the city of Arecibo, Puerto Rico. The radar consists of the world's largest dish antenna (305 m) with a powerful transmitter having 2.5 MW peak power. In combination with the reduced receiver noise temperature of ~ 75 K, this results in an overall system sensitivity more than 100 times greater than that of most other ISR facilities. The radar operates at 430 MHz.

5.1.1 Multiple radar auto-correlation function

The standard mode of the multiple radar auto-correlation function (MRACF) consists of simultaneous transmissions of multiple frequencies within a 250 kHz bandwidth. The technique itself has been developed by *Sulzer* [1986] to take full advantage of the high SNR achievable in Arecibo middle F-region measurements. It can be shown that once SNR surpasses 1, its further increase leads to little reduction in statistical errors in measurements. In such cases, by dividing the transmitting power among several radar channels, the excess signal can in effect be used to obtain additional statistically independent samples. In fact, some of the signal strength is traded for more independent samples or shorter integration time.

In the standard configuration, a 308 μs pulse is transmitted every 10 ms, and the backscatter is sampled in quadrature every 4 μs , producing a 250 kHz bandwidth. The transmitted pulse has been designed such that most of its energy appears at seven distinct peaks at seven frequencies with amplitudes of the same order of magnitude. The code generating the multiple frequencies is 28 μs in length (shown in Figure 5.1 along with its power spectrum), and the transmitted pulse is obtained by 11 repetitions of it. If we let $\phi(t)$ denote the basic code, the transmitted pulse can be written as

$$V_0(t) = \sum_{i=1}^{11} \phi(t - iT_b) \quad (5.1)$$

where $T_b = 28 \mu\text{s}$. The transmission and reception of the signal on each of the frequencies can be thought of as a separate long-pulse experiment, with 28 μs lag resolution. The altitude range in which the MRACF is normally used extends from 160 to 680 km. The range-gates are calculated every 38 km, and the data are typically integrated over 20 s intervals.

The procedure for choosing an appropriate code given a fixed number of spectral lines, n_{lines} , and fixed spectral spacing, f_{sp} , is presented in *Sulzer* [1986]. The inverse of the spectral spacing, t_{rep} , is the time that the code must occupy, and the ratio t_{rep}/n_{lines} gives the baud length for the code.

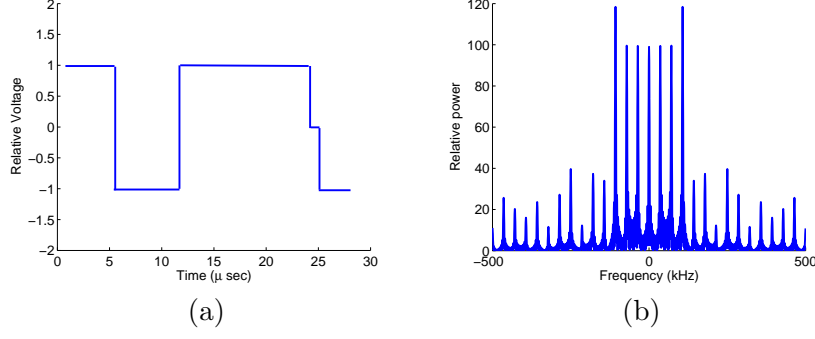


Figure 5.1 The basic code for generating the seven-frequency spectrum (a) and its spectrum (b).

5.2 Modified MRACF experiments at Arecibo

In this section, we describe the configuration of our experiments which utilize amplitude modulation to improve range resolution and statistical accuracy. In these new experiments, the basic transmitted pulse has been modified to transmit eight frequencies instead of seven in the standard mode. The procedure to search for an appropriate code to generate the eight-frequency spectrum is very similar to the one explained in [Sulzer, 1986]. Such a basic code is shown in Figure 5.2. In order to implement any coding scheme, we then need to modulate the basic code according to the desired waveform. To make it an eight-frequency spectrum, we repeat the code while flipping each baud sign after each repetition. In effect, the spacing is 50 KHz rather than 100 KHz, but the symmetry cancels half the lines so that we have eight lines spaced 100 KHz, and two weaker ones on the outside that we will not use. To be able to receive signals at all eight frequencies, we set the receiver bandwidth to 1 MHz. If we let $\phi(t)$ denote the basic code and $\{a_i\}_{i=1}^n$ represent the desired code sequence, then the modulated waveform can be expressed by

$$V_0(t) = \sum_{i=1}^n (-1)^{i-1} a(i) \phi(t - iT_b) \quad (5.2)$$

where T_b represents the length of the basic code (10 μ s).

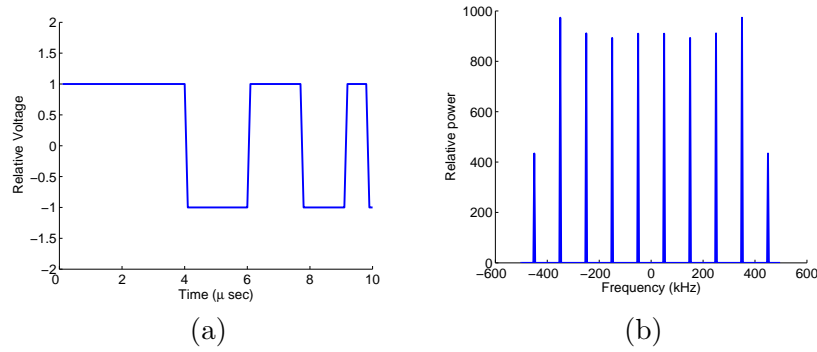


Figure 5.2 The basic code for generating the eight-frequency spectrum (a) and its spectrum (b).

5.3 Observations

5.3.1 July 2006 experiment

We start with presenting the results of observations on July 2 and 3, 2006, where a 13-bit Barker code, a long pulse, and three amplitude modulated codes were interleaved with a duration of 10 s. These codes shown in Figure 4.2 and denoted by AM1, AM2, and AM3 have on-off ratios of 0.75, 0.64, and 0.57, respectively. The data was taken using the feedline in the vertical position and the transmitter power was 1.5 MW. The observation started on July 2 at 19:00 LT and ended at 7:00 LT the next day. The 10.7 cm solar flux was between 85-87 throughout this time interval.

After performing the necessary data preprocessing, such as DC bias removal, etc., we computed the measured signal ACF from the raw data samples directly in the time domain. Background noise was estimated from the return from a few highest altitudes and was subtracted from the measured zero lag estimate prior to scaling and normalization. Note that the data was integrated over 10,000 pulse transmissions or 100 s.

Prior to ionospheric parameter estimation, let us compare the ACF obtained from different coding schemes. The first column in Figure 5.3 shows normalized long-pulse measured ACFs versus altitude in the form of error bars. The error bars themselves reflect auto-covariances calculated according to (4.7) (see [Lehtinen and Huuskonen, 1996; Hysell, 2000]). The second, third, and fourth columns show the altitude profiles of the measured ACFs using the three amplitude modulations (AM1 to AM3, respectively). Note that the error levels are substantially lower between 250 and 650 km for amplitude modulation schemes. Outside this range, however, error bars increase significantly due

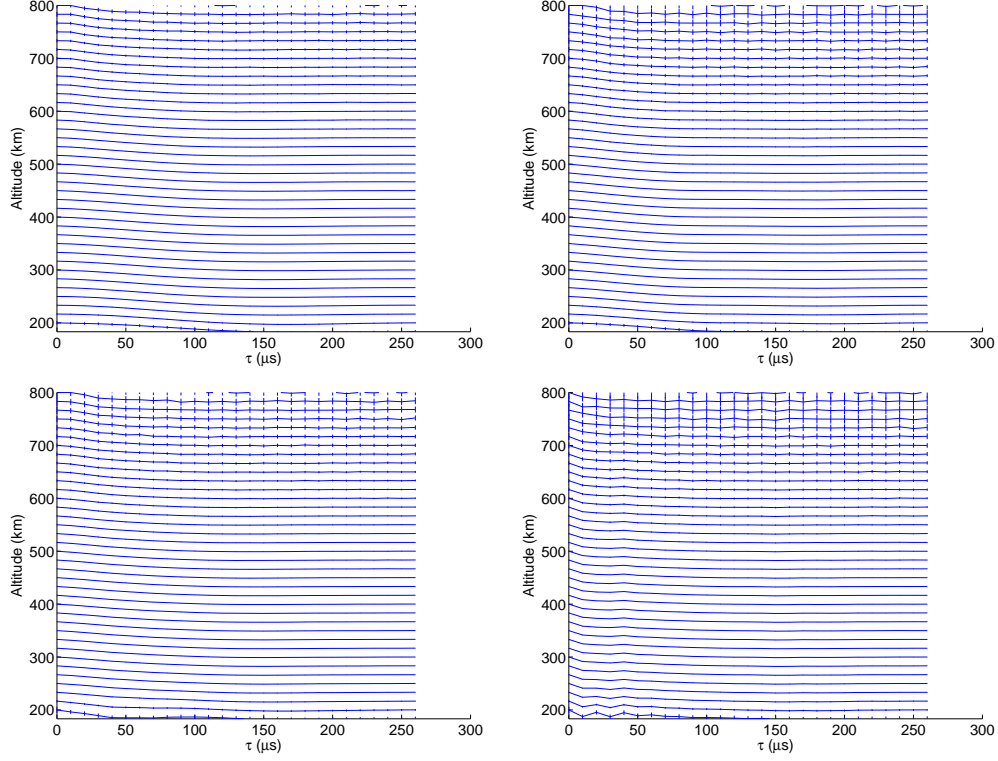


Figure 5.3 Altitude profiles of the measured ACF with the corresponding error bars for long-pulse and AM1 to AM3 from top left to bottom right.

to higher background noise levels. In fact, as expected, the lower the on-off ratio is for an amplitude modulation, the higher are the error levels. This also can be verified from the plot in Figure 5.4 where SNR is shown as a function of altitude for this particular data set between 20:04, and 20:17 LT. Above 600 km, the SNR reaches below 0.02, which is an indication of higher noise levels. Even with the same range resolution for all modulation schemes, we expect that amplitude modulation results in higher uncertainty levels on estimated parameters in such low SNR regimes.

To obtain altitude profiles of ionospheric parameters, we use the improved version of the hybrid technique, which increases the speed of analysis and reduces the memory requirements (see Section 4.2). In this method, we chose a grid for altitude profiles of ionospheric parameters over which the parameters are independent from one height to the next based on the altitude profile of SNR and the information of Table 4.1. Note that the grid that is typically used in the hybrid technique has a width equal to lag spacing ($10 \mu s \times c/2 = 1.5$ km, in this case). With such a grid, regularization is necessary as explained previously, and the trace of the influence matrix corresponding to the

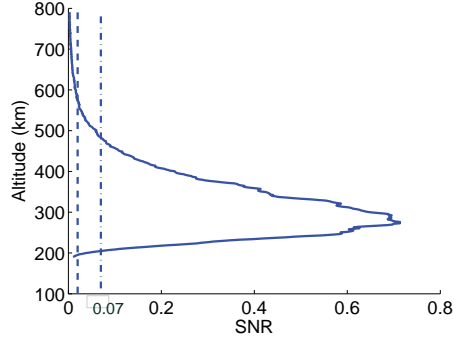


Figure 5.4 SNR as a function of altitude between 20:04 and 20:17 LT. The dotted-dash and dashed lines correspond to SNR reaching levels of 0.07, and 0.02, respectively.

optimal regularization parameter provides the range resolution or spacing between altitudes whose parameters are independent. If the new grid with independent altitudes (whose spacing is equal to the range resolution) is chosen for analysis, no regularization will be necessary, and the ACF lag estimates are obtained from a simple least-squares technique. This method not only reduces the computation time, but also reduces the memory which is required for the storage of the convolution matrix, error covariance, estimated lag profiles, and ultimately ionospheric parameters.

Because the optimal range resolution varies as a function of SNR, and SNR varies significantly through the range, we conclude that the optimal discretization grid for each modulation also varies as a function of height. In this work, however, for simplicity, we used the optimal range resolution corresponding to the high SNR for the whole altitude interval between 200 and 700 km. For instance, for the long pulse and AM3, discretization grids with spacings of 22.5 km and 9 km were used, respectively.

Once the discretization grid was chosen, the convolution matrix was constructed, and the ionospheric ACF lag estimates were computed from (3.9), where λ is set to zero and \mathbf{A} is the newly constructed convolution matrix. When this linear least-squares solution was obtained, the parameters were estimated via the nonlinear least-squares technique. Note that this problem belongs to the class of constrained optimization, as there are a few constraints associated with the physical parameters. For example, electron temperature T_e is never less than ion temperature T_i , or equivalently, $\mu = \frac{T_e}{T_i} \geq 1$. Another constraint is pertinent to ion fraction, p , which varies between 0 and 1, or $0 < p < 1$. To solve this bounded constrained nonlinear optimization problem, a trust region

algorithm is used [Coleman and Li, 1996]. For the optimization, we include ion temperature, T_i , electron-to-ion temperature ratio, μ , electron density, N_e and hydrogen fraction, p . Although μ is expected to be one during the night, we include it in the optimization as the observation continues until 7:00 LT the next morning.

Figure 5.5 (a)-(d) shows the estimated ion temperature for the long-pulse, AM1, AM2, and AM3 modulations, respectively, along with the error bars. The range resolution is the coarsest for the long-pulse modulation and the finest for AM3, which has the lowest on-off ratio. Although the range resolution is finer for amplitude modulated waveforms, the mean altitude profiles of estimated T_i (denoted by ‘o’) are nearly the same for all modulations. In addition, above 200 km and below 485 km, the error levels (denoted by the solid line) are also nearly equal. As the altitude increases above 485 km, we notice higher estimation errors for AM3, and as it further increases, higher error levels for AM2 and AM1. Note that SNR reaches below 0.07 at 485 km. Based on the information in Table 4.1 for AM3, the optimal range resolution which results in minimal estimation errors increases to $0.3 \times$ pulse length, i.e. 14.1 km, rather than 9 km which we use in our current analysis.

Figures 5.6, 5.7, and 5.8 (a)-(d) show the estimated electron-to-ion temperature ratio, normalized electron density, and hydrogen fraction as a function of altitude, respectively. Similarly to the Figure 5.5, we notice similar error levels below 485 km despite the finer resolution grids used in analysis of amplitude modulated data. Above this altitude, however, estimation errors increase dramatically as those resolution grids cannot support the data because of very low SNR (below 0.07). In Figure 5.7 (a)-(d) the normalized electron density obtained from the Barker code transmission is also included. Since the estimation result from the Barker code transmission is not dependent on the other modulations or analysis techniques that we have used in this experiment, the Barker code estimated electron density can provide a ground truth for our estimation results for different modulations. Note that the estimated electron density profiles from all modulations match the Barker code electron density profile closely.

Figure 5.9 shows the altitude profiles of SNR for the whole observation from 18:45 LT on July 2, 2006 to 7:08 LT on July 3, 2006. Figures 5.10, 5.11, 5.12, and 5.13 (a)-(d) represent the estimated ion temperature, temperature ratio, normalized electron density, and hydrogen fraction, respectively, for the whole time interval. Note that although discretization rids used for the the long

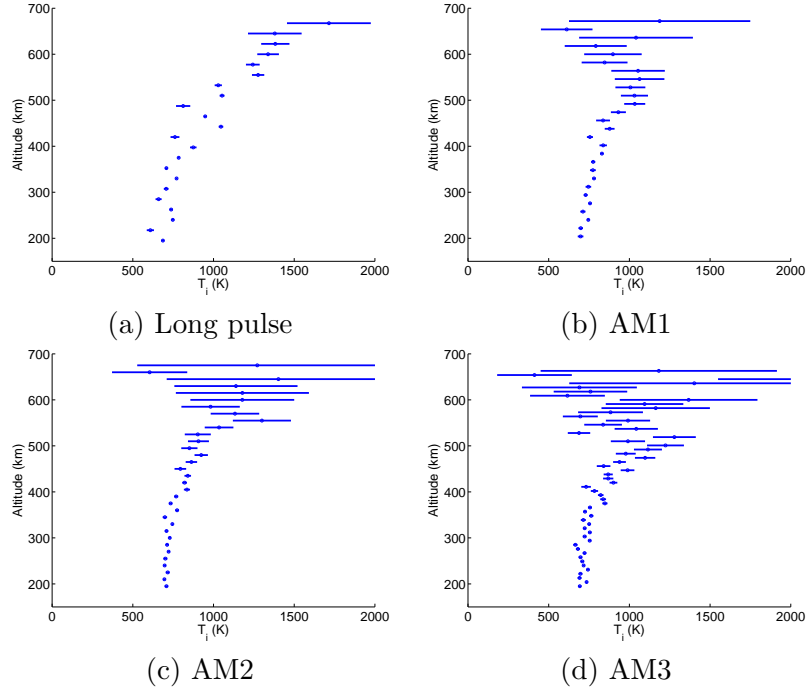


Figure 5.5 Altitude profiles of estimated ion temperature for (a) long-pulse, (b) AM1, (c) AM2, and (d) AM3 with 100 s integration time between 20:04 and 20:17 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively. The altitude profiles of the mean estimated parameters are nearly the same, so are the error levels between 200 km and 485 km, for all modulation schemes. The error levels increase more significantly for amplitude modulation with lower on-off ratios above 485 km where SNR reaches below 0.07.

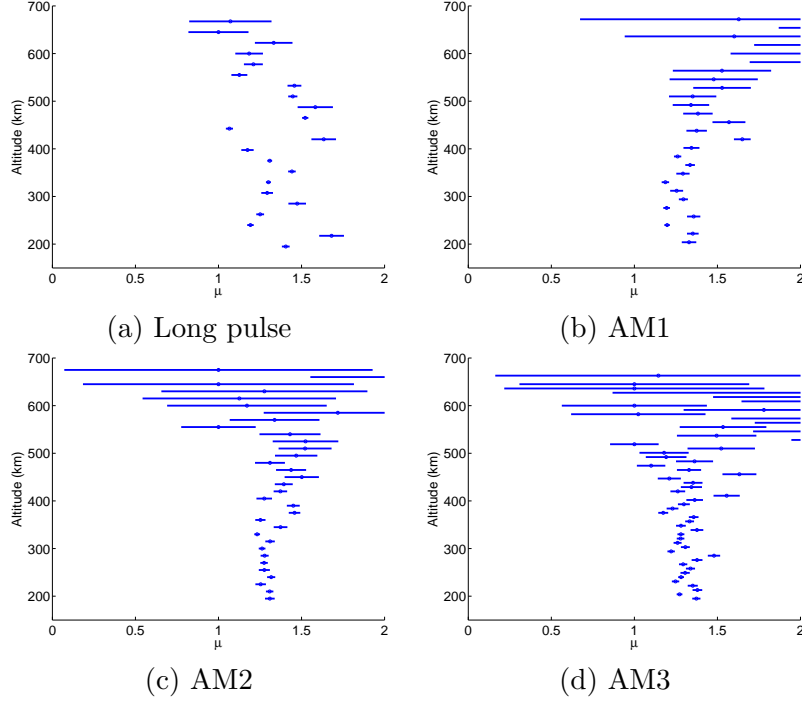


Figure 5.6 Altitude profiles of estimated electron-to-ion temperature ratio for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3 with 100 s integration time between 20:04 and 20:17 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively.

pulse, AM1, AM2, and AM3 were different (22.5 km, 17.1 km, 13 km, and 9 km, respectively) the mean estimated parameters are nearly equal for all modulation schemes. Note also that we expect nearly similar errors for estimation results from all modulations with higher errors for amplitude modulation (especially with lower on-off ratios) when SNR reaches below 0.07. The images in Figures 5.14, 5.15, 5.16, and 5.17 (a)-(d), which represent the estimation errors for each parameter, verify this expectation.

5.3.2 August 2005 experiment

Next, we present the estimation results from an observation conducted with the Arecibo radar on August 14, 2005 from 20:00 LT to 4:00 LT the next day. The purpose of this experiment was not only to compare the performance of long-pulse modulation versus amplitude modulation schemes with various on-off ratios but also to compare the performance of several amplitude modulated pulses with the same on-off ratios. The general configuration of the experiment is described at the beginning of Section 5.2; i.e. the basic code that generates the eight-frequency spectrum is used. In

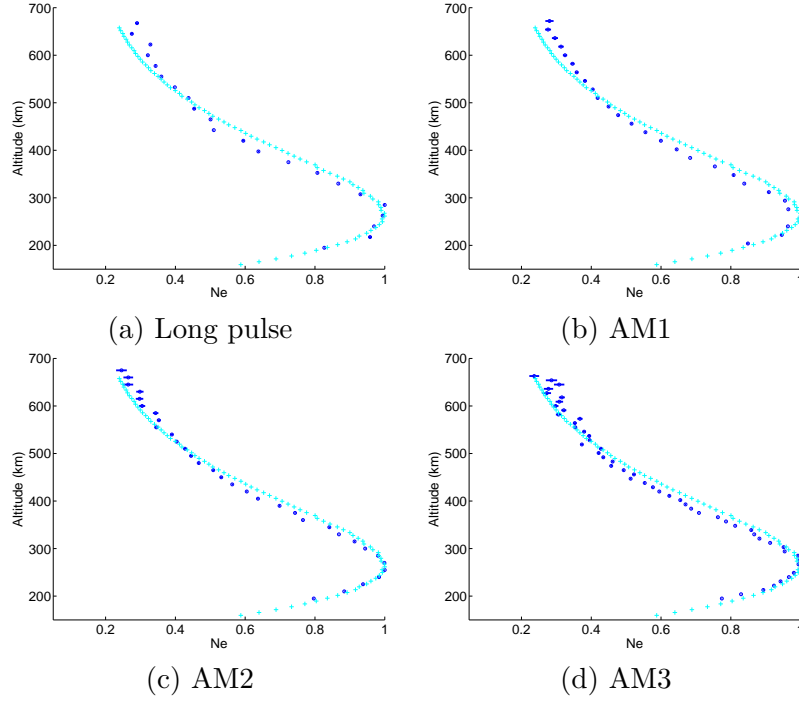


Figure 5.7 Altitude profiles of estimated electron density for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3 with 100 s integration time between 20:04 and 20:17 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively. The electron density profile obtained from the Barker codes is also included in the plot (denoted by +). Note that the estimated electron density profile from all modulations match the Barker code electron density profile closely.

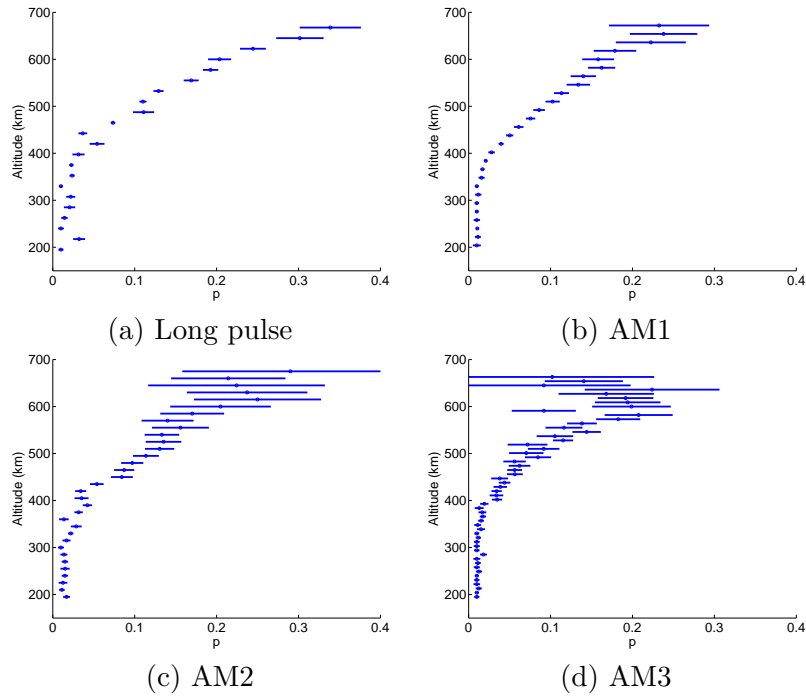


Figure 5.8 Altitude profiles of estimated hydrogen fraction for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3 with 100 s integration time between 20:04 and 20:17 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively.

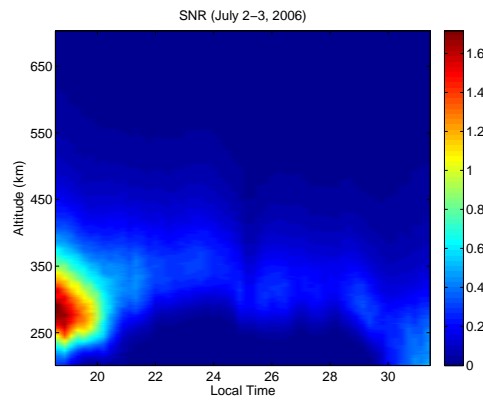


Figure 5.9 SNR as a function of altitude for the whole observation time from 18:45 LT on July 2 2006 to 7:08 LT the next day.

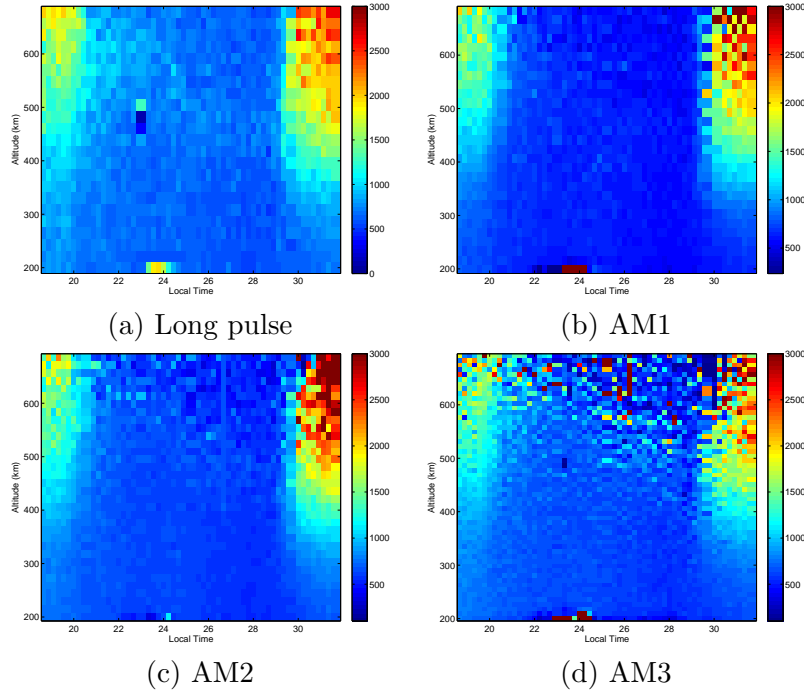


Figure 5.10 Estimated mean ion temperature as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

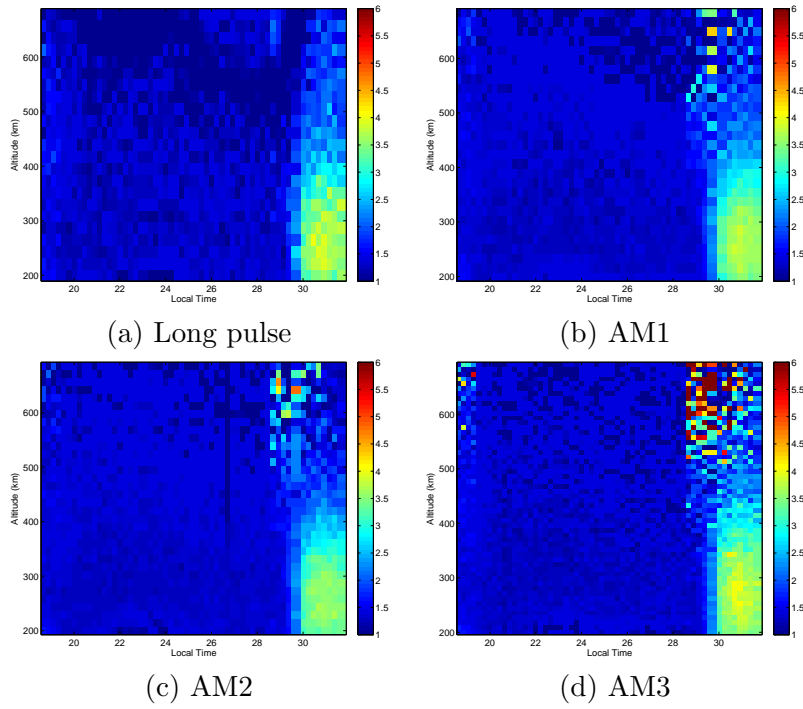


Figure 5.11 Estimated mean electron-to-ion temperature ratio as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

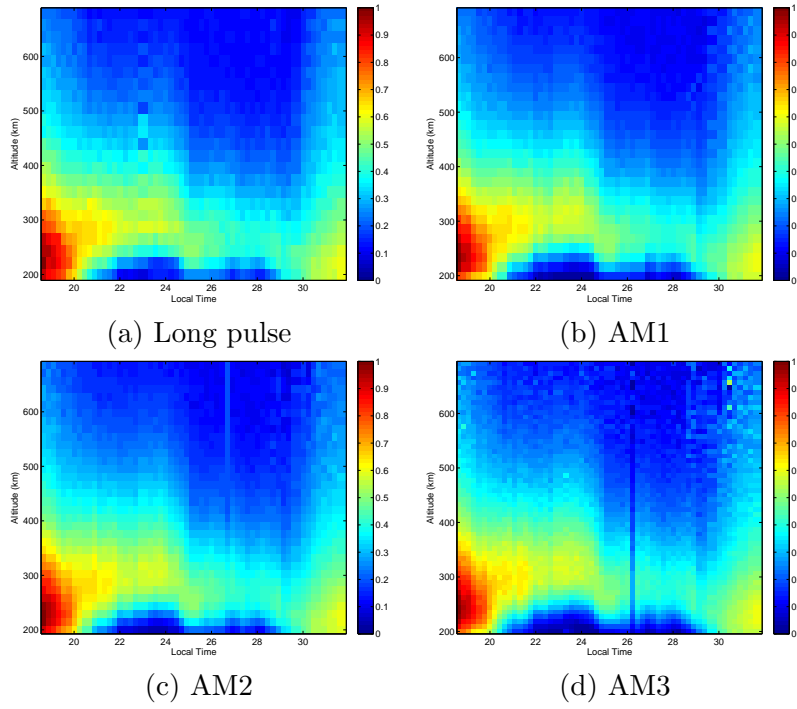


Figure 5.12 Estimated mean electron density (normalized) as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

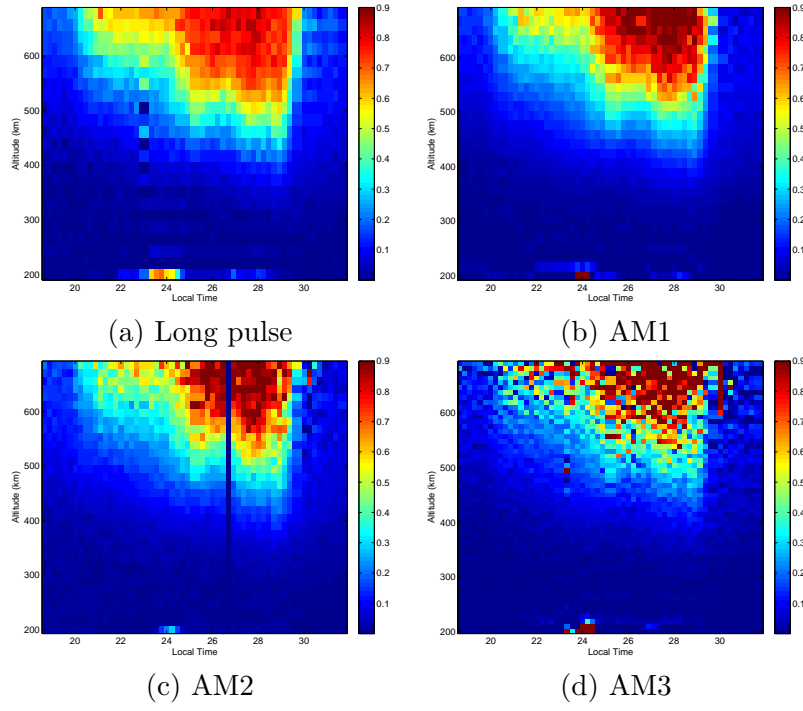


Figure 5.13 Estimated mean hydrogen fraction as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

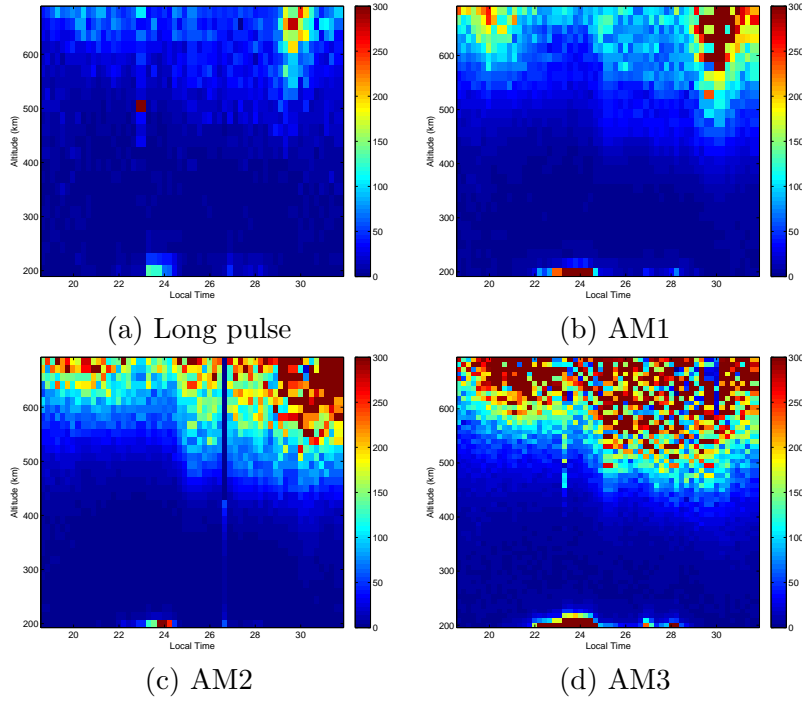


Figure 5.14 Estimated errors in ion temperature as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

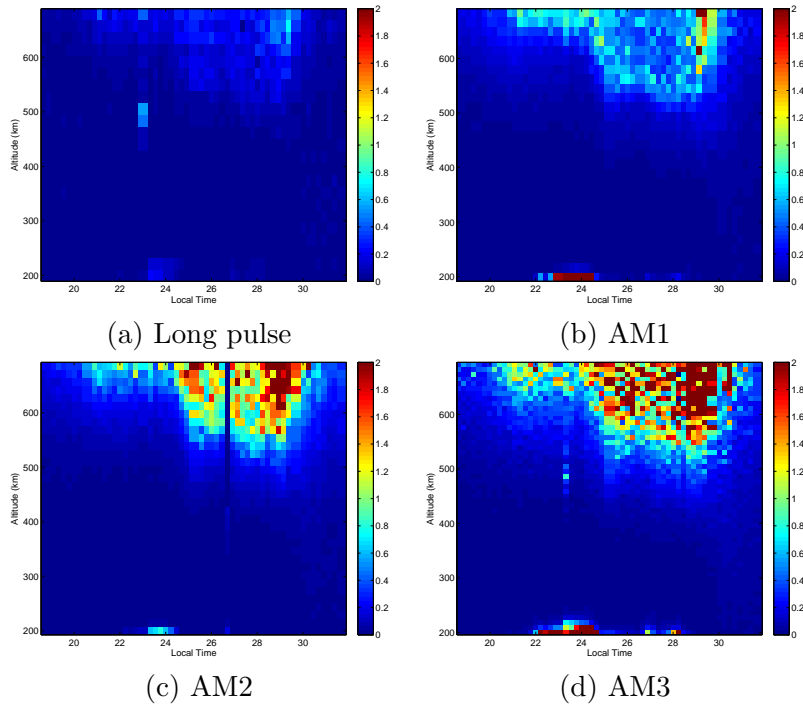


Figure 5.15 Estimated errors in electron-to-ion temperature ratio as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

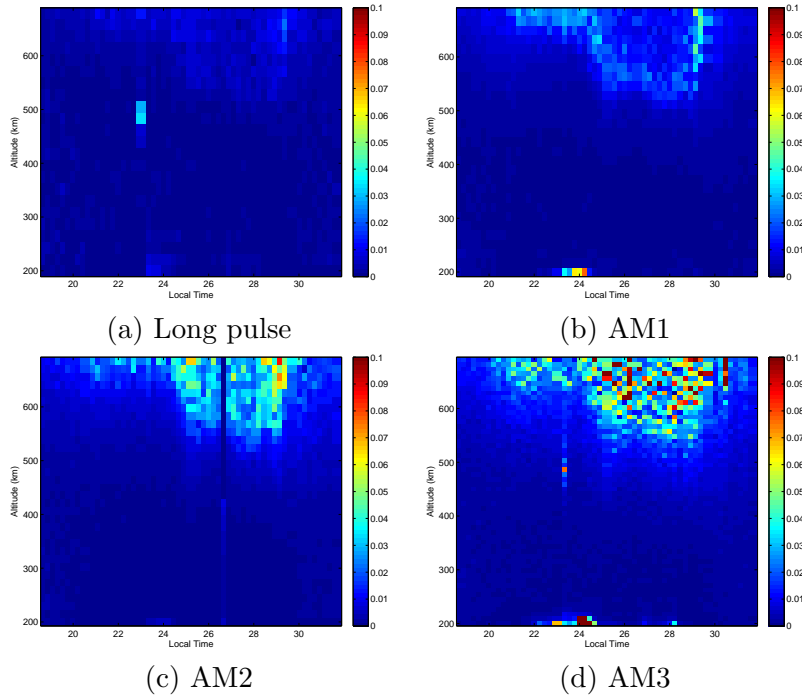


Figure 5.16 Estimated errors in electron density as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

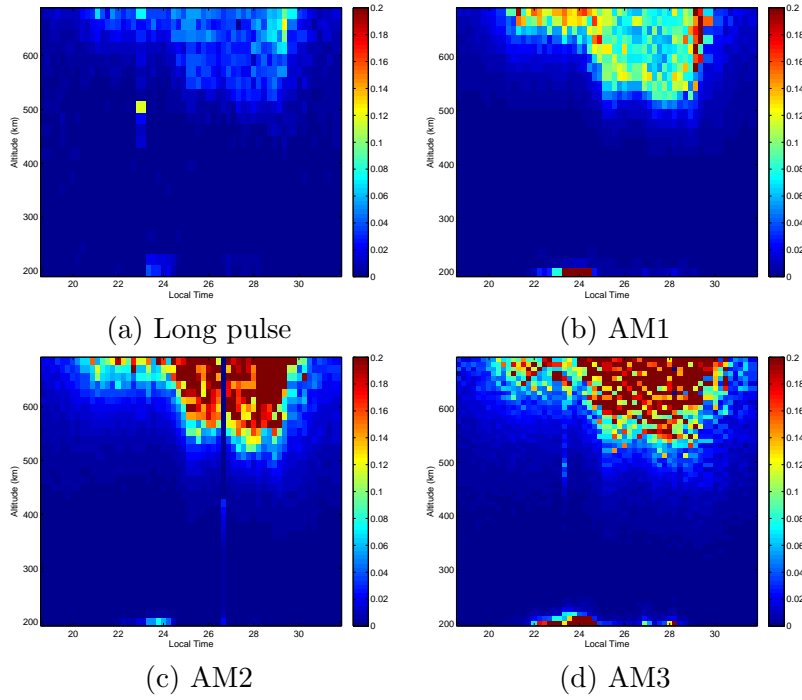


Figure 5.17 Estimated errors in hydrogen fraction as a function of altitude and time between 18:45 LT on July 2, 2006 and 7:08 LT on July 3, 2006 for (a) long pulse, (b) AM1, (c) AM2, and (d) AM3. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

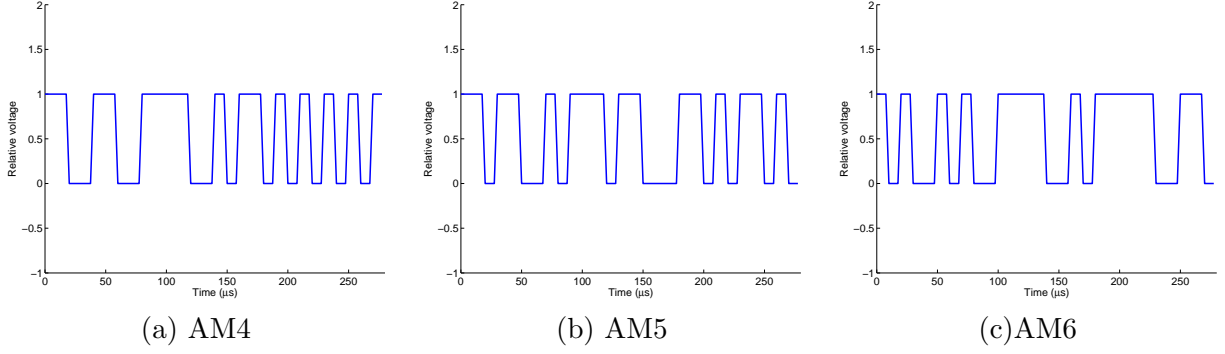


Figure 5.18 Specific amplitude modulations used for the modified MRACF experiments between August 14 and 15, 2005. The other four modulations utilized in this experiment were long-pulse, and AM1 to AM3 shown in Figure 4.2.

this experiment, a long pulse and six amplitude-modulated codes were interleaved with a duration of 10 s. Three of these codes, AM1, AM2, and AM3, are shown in Figure 4.2, with on-off ratios of 0.75, 0.64, and 0.57, respectively. The other three AM pulses denoted by AM4, AM5, and AM6 have also on-off ratios of 0.57, and are shown in Figure 5.18.

For the analysis of long-pulse, AM1, AM2, AM3, AM4, AM5, and AM6 measurements, discretization grids of 22.5 km, 17.1 km, 13 km, 9 km, 9 km, 9 km, and 9 km were used, respectively. These range resolutions correspond to the optimal resolution for high SNR. Therefore, again we expect similar estimation errors for all modulations when the SNR is high. The data are integrated over 8000 pulse transmissions, or 80 s. Figure 5.19 shows the altitude profile of SNR as a function of time for the whole observation time. The analysis technique was very similar to the one used for the 2006 experiment. The parameters included in the nonlinear optimization were ion temperature, electron density, and hydrogen ion fraction. Because it was mainly a nighttime experiment, we set the electron-to-ion temperature ratio to be one as a side constraint. Other bound constraints mentioned previously also remained in effect.

Figures 5.20, 5.21, and 5.22 (a)-(g) show the estimated ion temperature, temperature ratio, normalized electron density, and hydrogen fraction, for long pulse, and AM1 to AM6, respectively, for the whole time interval. Note that the mean estimated parameters are nearly equal for all modulation schemes. Also, we observe nearly similar errors for estimation results from all modulations for $\text{SNR} > 0.07$, and higher errors for amplitude modulation (especially for the ones with lower on-off ratios) when $\text{SNR} < 0.07$ (see the images in Figures 5.23, 5.24, and 5.25 (a)-(g) representing the

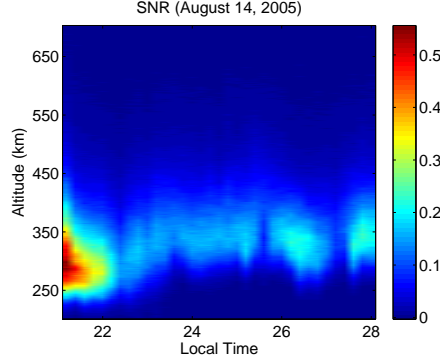


Figure 5.19 SNR as a function of altitude between 20:00 LT on August 14, 2005, and 4:00 LT the next day.

estimation errors for each parameter).

Because all of the AM3 to AM6 modulations have the same on-off ratio (0.57), they yield similar estimation errors as AM1 to AM3 up to 450 km and higher errors above 450 km. However, AM5 results in slightly smaller estimation errors (especially in electron density and hydrogen fraction), and the estimated mean electron density from AM5 measurements shows smaller variations at the high altitude range (about 600 km). The reason for this slight difference in performance is that although they have similar on-off ratios, they differ in other characteristics. In other words, the on-off ratio of a code provides only one measure for comparison. In order to develop the optimal coding scheme, we need to consider other characteristics of amplitude modulated codes, such as the condition number of associated convolution matrices which provides a criterion for conditioning of the inverse problem.

5.3.3 April 2004 experiment

This experiment is different from the other two experiments in that the observations were made during daytime from 14:50 to 17:26 LT rather than at night. Since the backscattered signal is much stronger in daytime compared with nighttime, due to higher electron density, the SNR is also higher, which makes amplitude modulation transmission more suitable. Comparison of the altitude profile of SNR at 14:50 LT, shown in Figure 5.26, with the SNR altitude profile at 20:04 LT, shown in Figure 5.4, indicates this huge difference. In this experiment, AM1 and AM2 were interleaved with durations of 10 s.

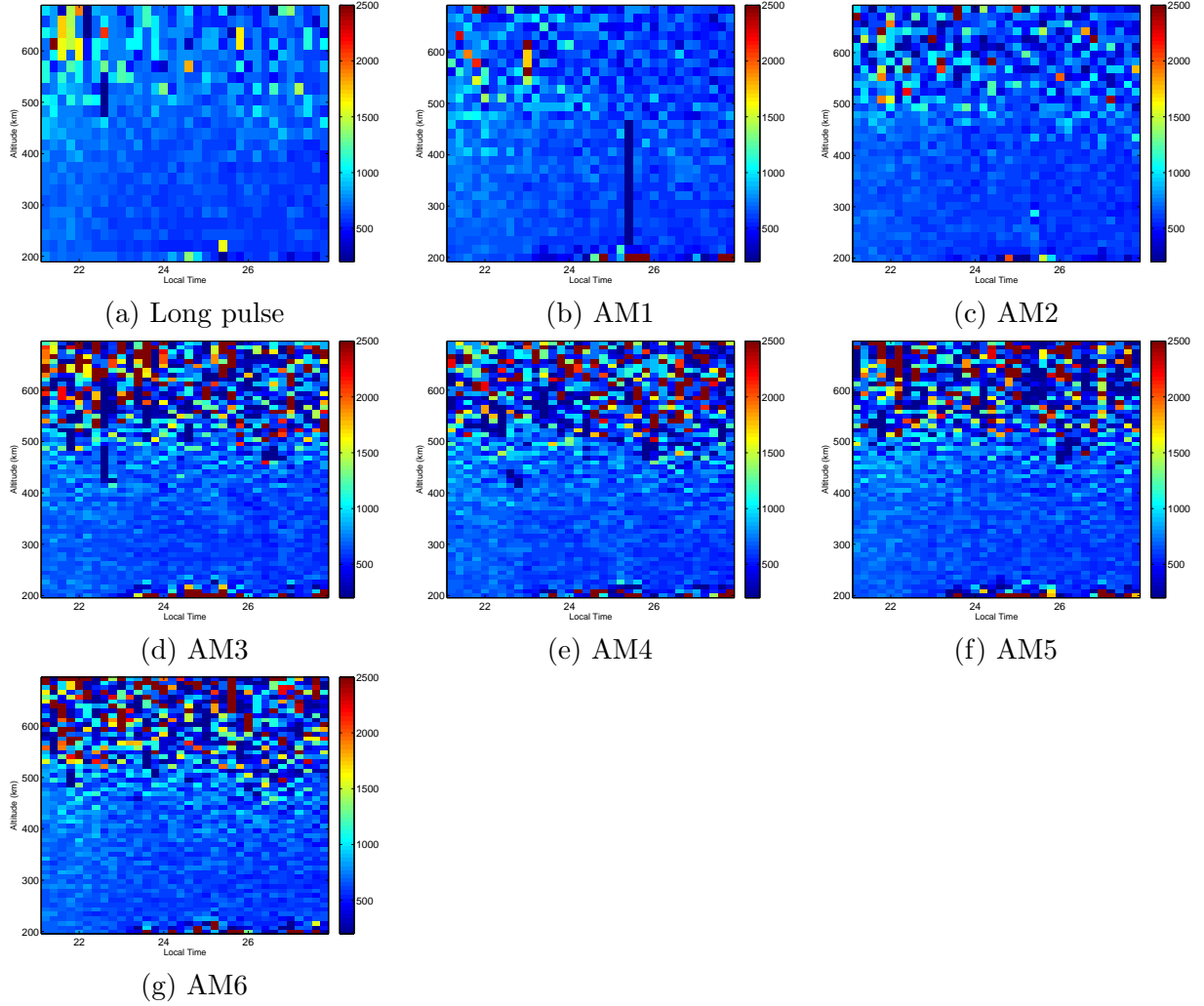


Figure 5.20 Estimated mean ion temperature as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

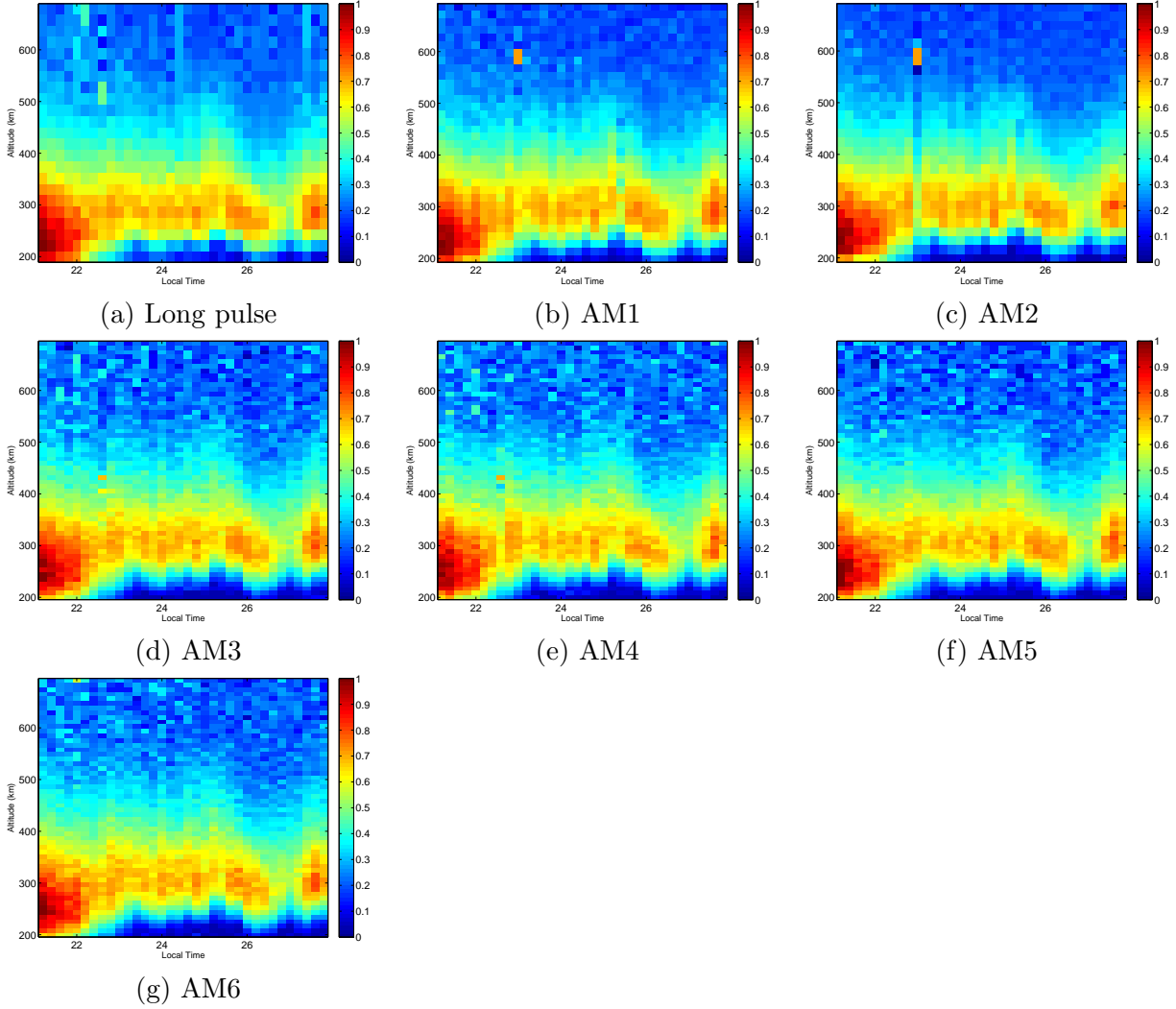


Figure 5.21 Estimated mean electron density as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

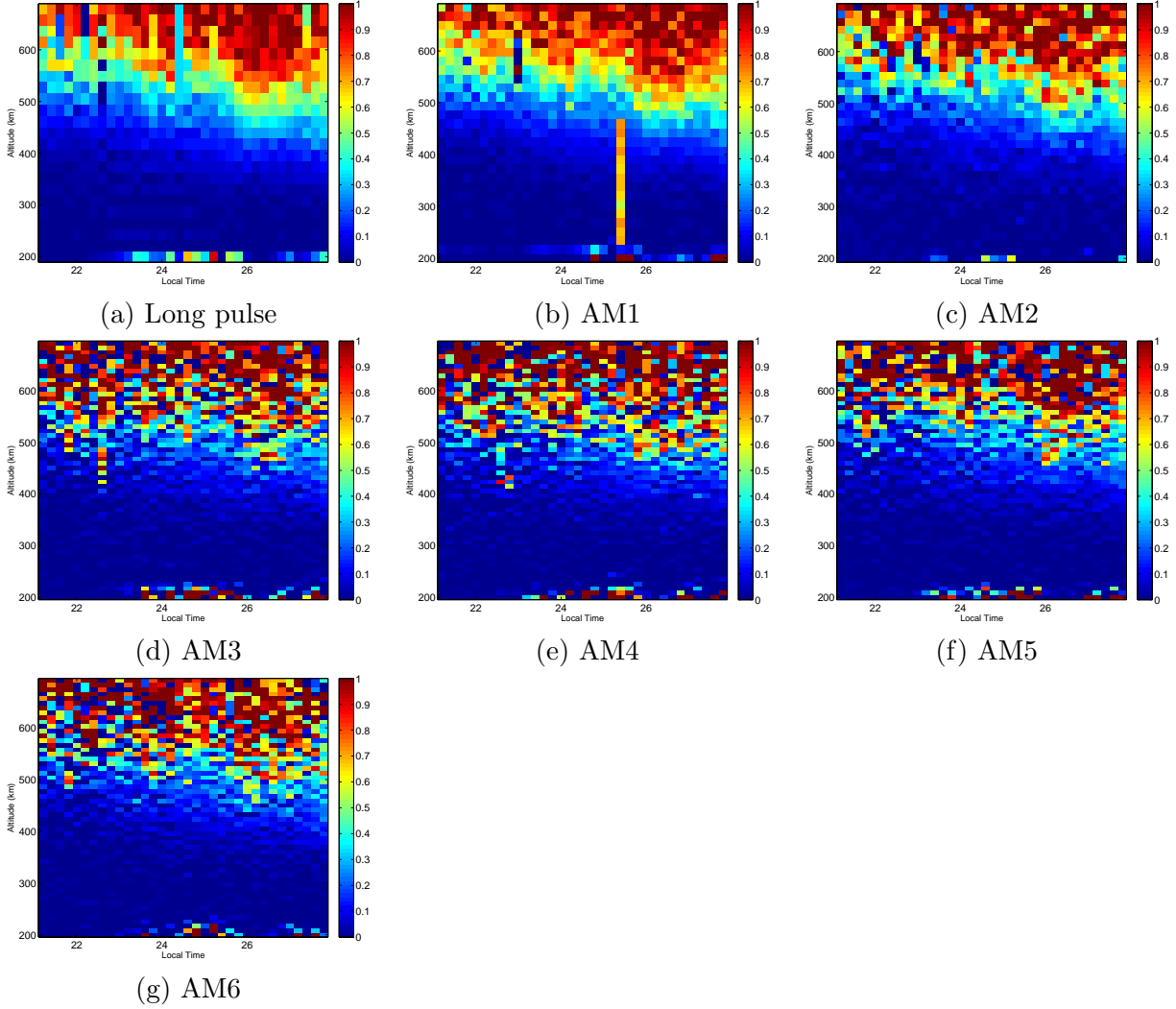


Figure 5.22 Estimated mean hydrogen fraction as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

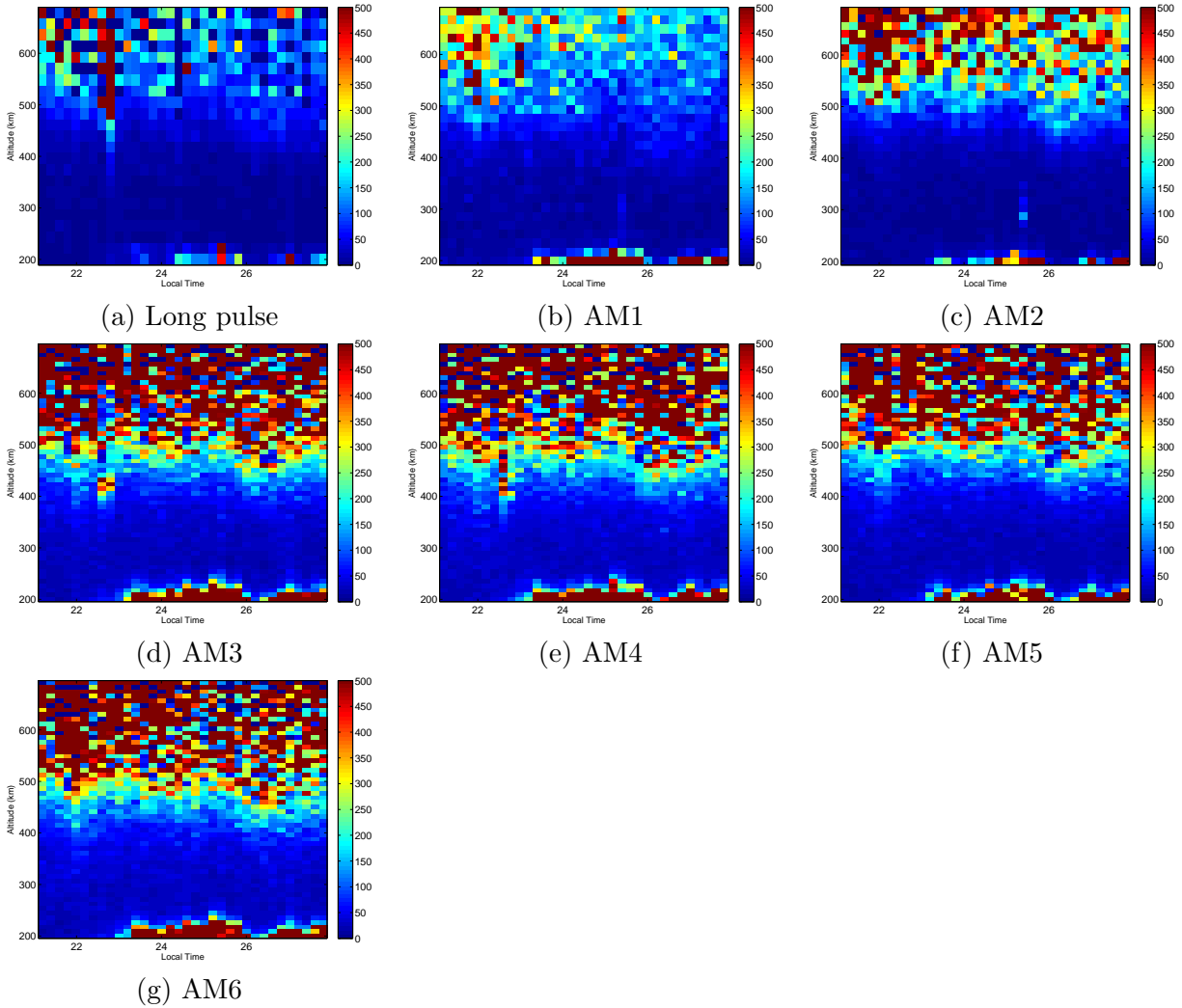


Figure 5.23 Estimated errors in ion temperature as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

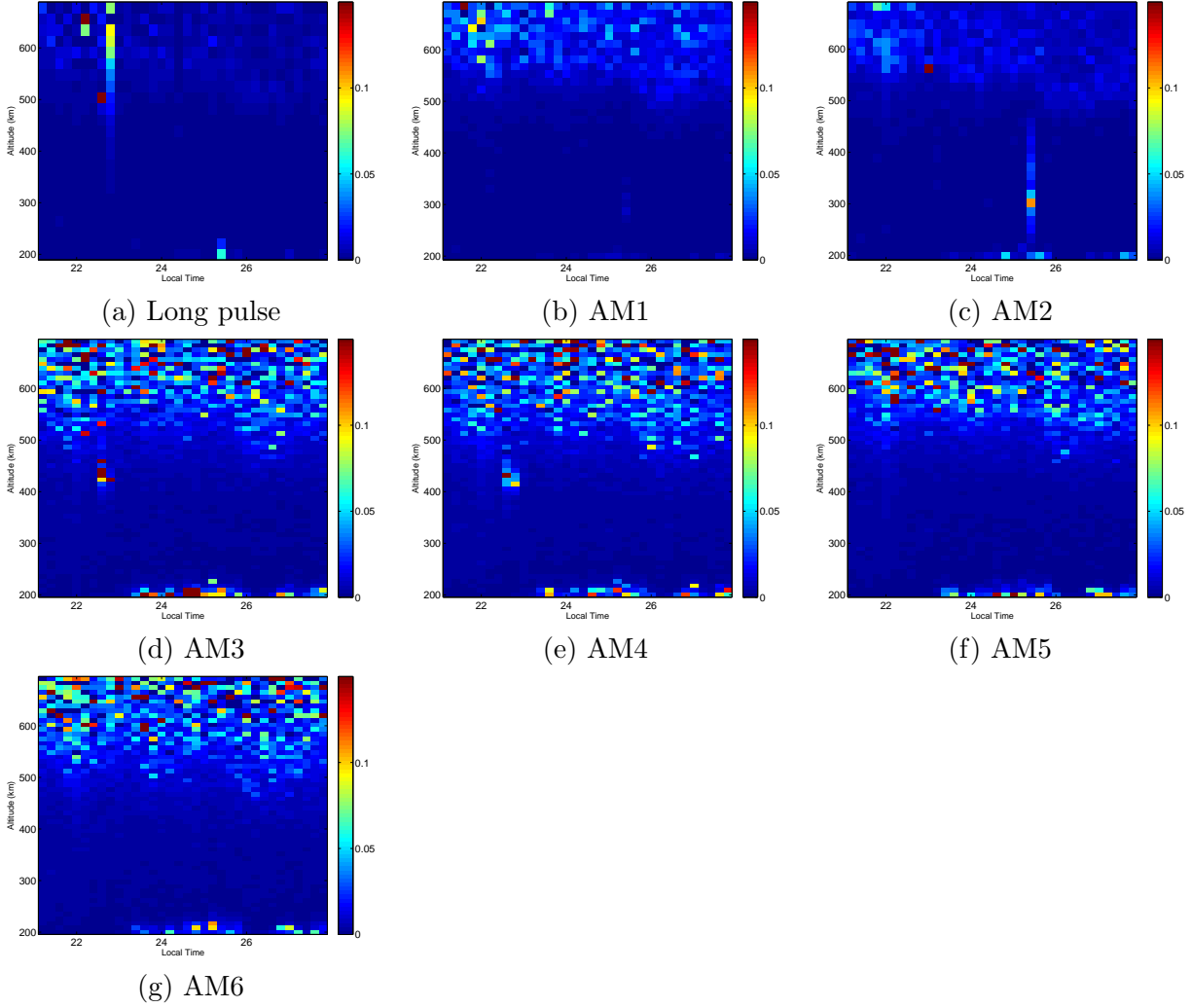


Figure 5.24 Estimated errors in electron density as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

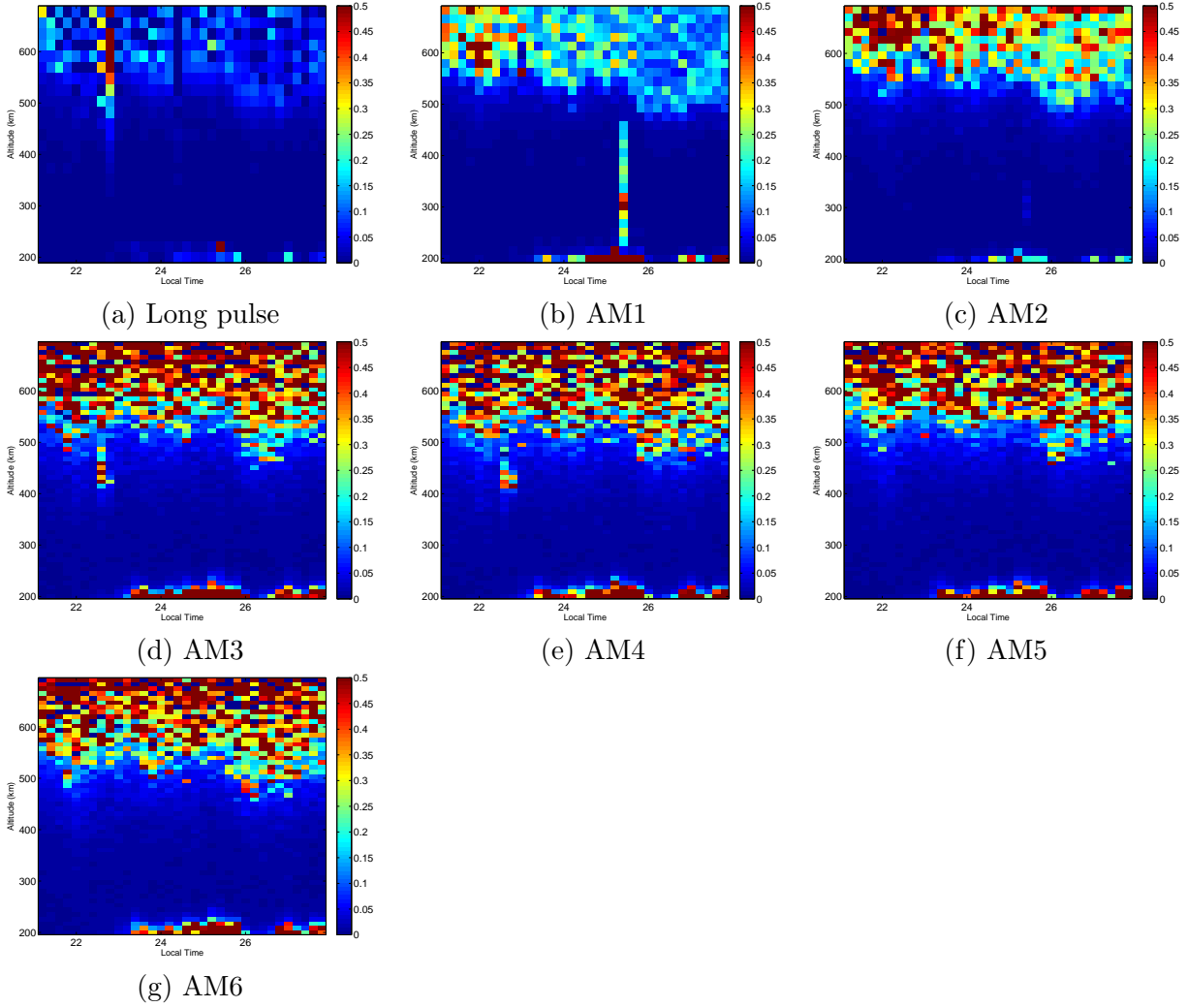


Figure 5.25 Estimated errors in hydrogen fraction as a function of altitude and time between 20:00 LT on August 14, 2005 and 4:00 LT on August 15, 2005 for (a) long pulse, (b) AM1, (c) AM2, (d) AM3, (e) AM4, (f) AM5, and (g) AM6. Note nearly similar error levels for all modulation for lower altitudes where SNR is higher (above 0.07) and higher error levels for amplitude modulations (especially those with lower on-off ratios) for altitudes where SNR reaches below 0.07.

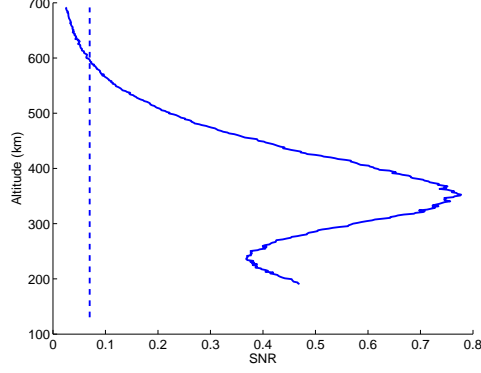


Figure 5.26 SNR as a function of altitude between 14:50 and 15:01 LT on April 3, 2004. The dashed line corresponds to SNR reaching the level of 0.07.

For analysis, discretization grids of lengths 17.1 km and 13 km were used for AM1 and AM2, respectively. For optimization we included the ion temperature, electron-to-ion temperature ratio and electron density. Figures 5.27 to 5.29 (a) and (b) show the mean estimated T_i , μ , and N_e for AM1 and AM2, respectively. The estimated mean parameters and the corresponding error bars are denoted by o's and solid lines, respectively. As can be seen from these figures, the mean estimated parameters as well as estimation errors are nearly the same for the whole range of altitudes from 200 km to 650 km.

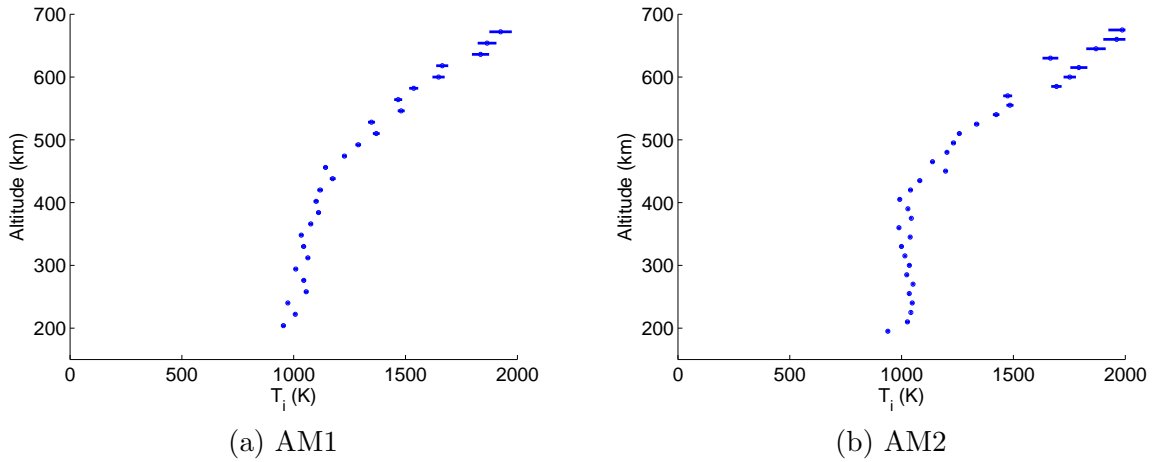


Figure 5.27 Altitude profiles of estimated ion temperature for (a) AM1 and (b) AM2 with 130 s integration time between 14:48 and 15:01 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively. The altitude profiles of the mean estimated parameters are nearly the same, so are the error levels between 200 km and 670 km, for both modulation schemes.

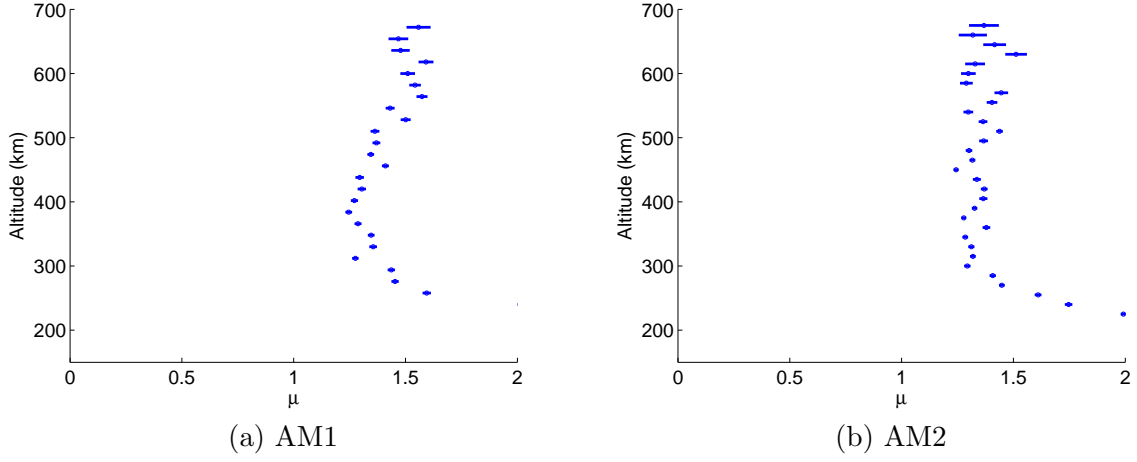


Figure 5.28 Altitude profiles of estimated electron-to-ion temperature ratio for (a) AM1 and (b) AM2 with 130 s integration time between 14:48 and 15:01 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively. The altitude profiles of the mean estimated parameters are nearly the same, so are the error levels between 200 km and 670 km, for both modulation schemes.

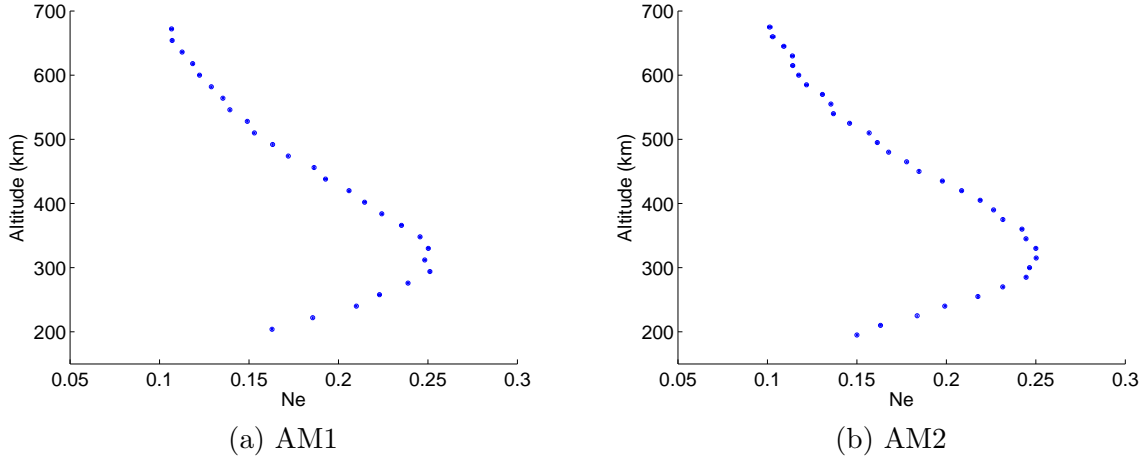


Figure 5.29 Altitude profiles of estimated electron density for (a) AM1 and (b) AM2 with 130 s integration time between 14:48 and 15:01 LT. The mean estimated values and estimation error are indicated by o's and solid lines, respectively. The altitude profiles of the mean estimated parameters are nearly the same, so are the error levels between 200 km and 670 km, for both modulation schemes.

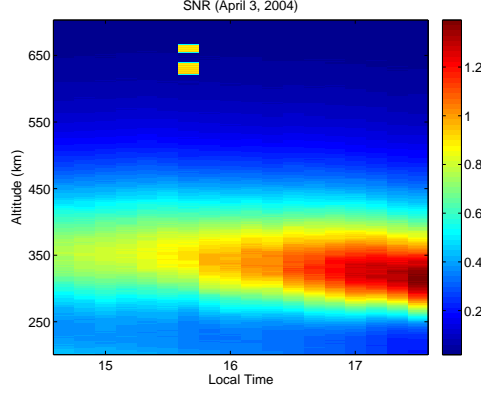


Figure 5.30 SNR as a function of altitude for the whole observation time from 14:48 LT on April 3, 2004 to 17:30 LT.

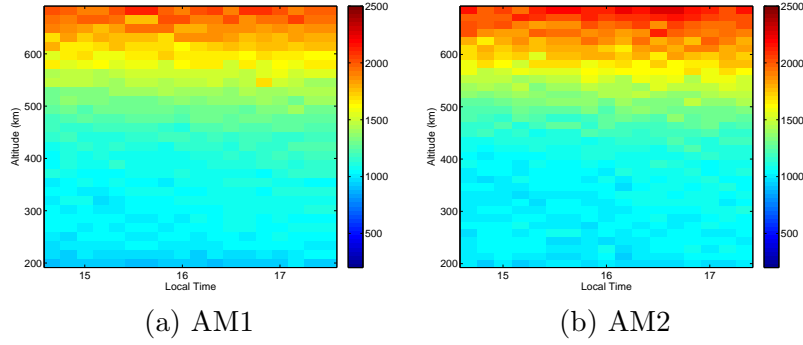


Figure 5.31 Estimated mean ion temperature as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

Figures 5.30 to 5.33 show the altitude profiles of SNR, T_i , μ , and N_e as function of time between 14:48 and 17:30 LT, respectively. Figures 5.34 to 5.36 illustrate the estimation errors corresponding to plasma parameters for the same time interval. Note that the error levels are nearly equal in both cases despite different discretization grids.

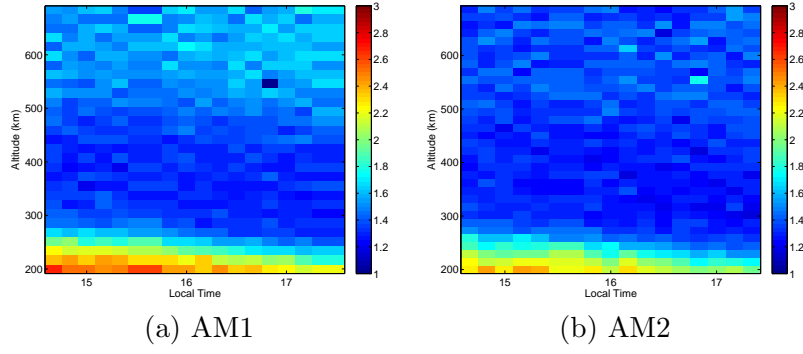


Figure 5.32 Estimated mean electron-to-ion temperature ratio as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

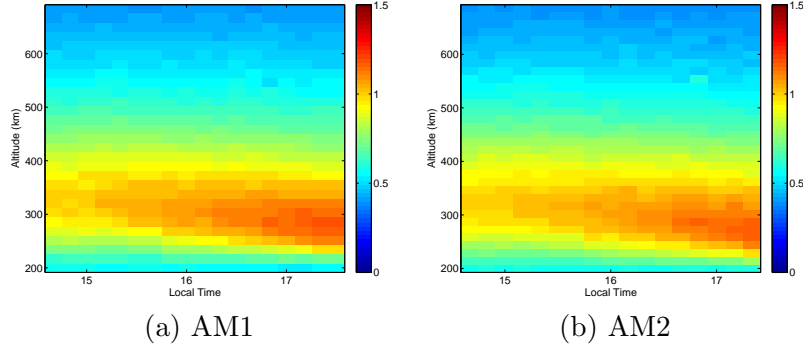


Figure 5.33 Estimated mean electron density as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. The estimated mean parameters are nearly equal for all modulations despite different range resolutions.

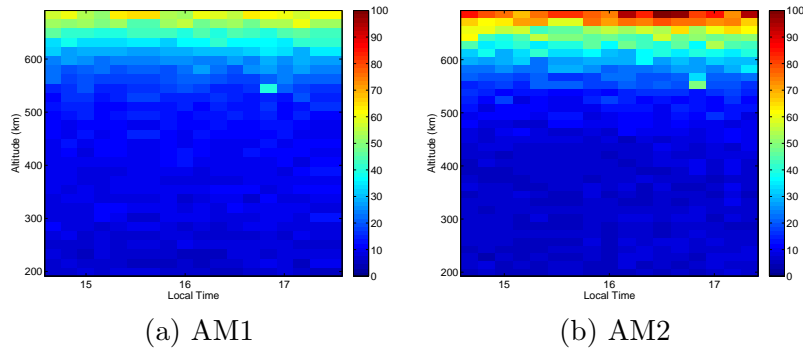


Figure 5.34 Estimated errors in ion temperature as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. Note that the error levels are nearly equal in both cases despite different discretization grids.

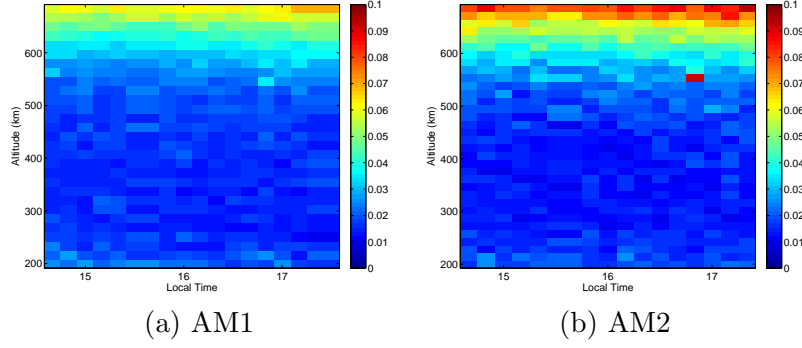


Figure 5.35 Estimated errors in electron-to-ion temperature ratio as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. Note that the error levels are nearly equal in both cases despite different discretization grids.

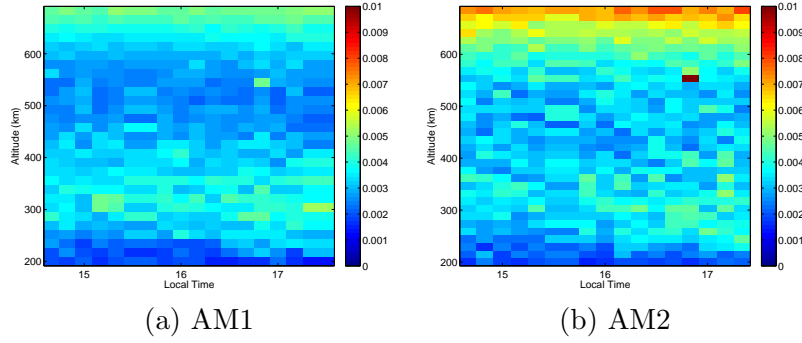


Figure 5.36 Estimated errors in electron density as a function of altitude and time between 14:48 and 17:30 LT on April 3, 2004 for (a) AM1 and (b) AM2. Note that the error levels are nearly equal in both cases despite different discretization grids.

CHAPTER 6

OPTIMAL MODULATION IN F-REGION INCOHERENT SCATTER EXPERIMENT DESIGN

Both numerical and experimental results presented previously indicate that amplitude modulation shows improvement over long-pulse modulation in terms of range resolution and statistical accuracy. This improvement is more pronounced during daytime when SNR is medium to high for AM pulses with smaller on-off ratios. However, slight differences between the performances of AM codes with the same on-off ratios necessitate consideration of other criteria for optimal modulation design. Such criteria include the sum of squared errors (SSE), uniformity of squared errors, and condition number of convolution matrices.

An AM code with an on-off ratio of $\frac{n_p}{N}$ can be viewed as a combination of n_p short pulses located at multiple integers of the baud length. With fixed baud and pulse lengths ($\Delta\tau$ and $\Delta\tau N$, respectively), the search space includes all AM codes that consist of n_p short pulses. The task is to search over a set of candidate locations for short pulses to find the optimal or close-to-optimal subset. Due to the vast search space, however, an exhaustive search is not computationally feasible. Therefore the problem of finding the optimal amplitude modulation should be viewed and tackled as an optimization problem.

In this chapter, we employ mathematical approaches currently used in observation selection to solve this optimization problem [Golub and Van Loan, 1989; Miller, 2002]. The task of modulation selection is viewed so as to choose the location of n_p short pulses such that an optimality criterion is met.

This chapter is organized as follows: we first provide an overview of the experiment design in view of model order selection methods. Next, we describe the statistical formulation and several

optimality criteria. Finally, we present the selection algorithm and investigate its performance through numerical experiments.

6.1 Experiment design in view of subset selection

In the ISR experiment design, we are interested in pulses that support the finest range resolution in estimated parameters or result in the smallest estimation errors (for a fixed resolution) for a fixed SNR. The results indicated in the previous two chapters demonstrated that the on-off ratio of an amplitude-modulated pulse serves as a determining factor. However, other factors, such as the configuration of the code itself (location of the short pulses in the code), are equally important, because they determine the condition number of convolution matrices, and the number of integrated heights for each ACF lag measurement. For the experiment design using amplitude modulation with an on-off ratio of $\frac{n_p}{N}$, there are N possible candidates for locations of n_p short pulses. For the optimal design, these locations should be selected such that the most accurate estimation of plasma ACF at individual altitudes and, ultimately the most accurate estimation of plasma parameters, is obtained.

An optimality criterion is needed for the choice of the n_p locations, and the resulting combinatorial optimization problem must be tackled. The optimality criterion is designed such that it only depends on the statistics of measurement errors Σ_ϵ , and the convolution matrix for each lag. We should mention that in our application of interest, especially in medium to high SNR scenarios, the error covariance matrix is affected by both the convolution matrix and the SNR, and that different pulse configurations result in different forward models. However, for a fixed SNR, we can form the following optimization problem [Sharif, 2004]

$$\mathcal{S}^* = \arg \min_{\mathcal{U} \subseteq \mathcal{S}, |\mathcal{U}|=n_p} \text{Cost}(\mathbf{A}^{(\mathcal{U})}) \quad (6.1)$$

where \mathcal{S} denotes the set of all candidate locations, and $\mathbf{A}^{(\mathcal{U})}$ represents the convolution matrix associated with $\mathcal{U} \subseteq \mathcal{S}$. The cardinality is denoted by $|\cdot|$, and $\text{Cost}(\mathbf{A}^{(\mathcal{U})})$ is the cost of choosing subset \mathcal{U} as the short-pulse locations. The best subset can be found by an exhaustive search over all

$\binom{N}{n_p}$ possible combinations. However, this grows exponentially in N , and is impractical even for a moderate number of candidate locations. Therefore, in order to make the search computationally feasible, several heuristics which do not require exhaustive searches have been proposed [Miller, 2002]. One such algorithm is the “greedy” algorithm of Golub and Van Loan [1989]. The greedy algorithm is a sequential backward selection scheme [Reeves and Zhe, 1999; Crouveur and Bresler, 2000]. The idea is to start off with short pulses at all N candidate locations (long pulse), and remove one short pulse at a time until n_p short pulses are present in the code. The short pulse that is removed should be chosen such that the optimal solution to (6.1) is obtained. We refer to this approach as the sequential backward selection (SBS) algorithm [Reeves and Zhe, 1999]. It is worth mentioning that there is also the forward greedy algorithm [Marill and Green, 1963] in which the best short pulse location is chosen initially, and at each iteration another short pulse is added until there are n_p short pulses present in the amplitude-modulated code. This method, however, is not suitable for our application because a single short pulse does not provide any lag information.

In this work we use a modified version of the SBS algorithm. Our problem is different from the observation selection problem in a sense that we cannot just eliminate one row at a time from the observation kernel [Reeves and Heck, 1993]. At every iteration, when a short pulse is removed from the subset, the configuration of the forward model, or convolution matrix, changes completely. At every iteration, the cost of removing a short pulse at a particular location is calculated, and the short pulse which yields the optimum cost is eliminated.

6.2 Statistical formulation

In the previous chapters, it has been well established that utilization of the long-pulse modulation results in an ill-posed inversion of ISR measurements. In order to overcome the ill-posed nature of the problem, we are required to apply regularization techniques. However, as shown in Chapter 4, if the resolution is set to fundamental resolution supported by measurements, for example half the pulse length for a long pulse, no regularization is necessary and the plasma ACF lags are obtained via the least-squares technique or the maximum likelihood (ML) approach (see Section 4.3). Therefore,

following the notation in Chapter 4, the ML estimate of the plasma ACF lags is represented by

$$\hat{\mathbf{m}}_{ML} = \arg \min_{\mathbf{m}} \|\mathbf{y} - \mathbf{A}\mathbf{m}\|_{\Sigma_{\epsilon}}^2 \quad (6.2)$$

$$= (\mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A})^{-1} \times \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{y} \quad (6.3)$$

where Σ_{ϵ} is the measurement error covariance matrix, and is obtained from (4.7) to (4.10). As shown in Section 4.1, the structure of the matrix is highly affected by the transmitted pulse envelope.

The above statistical framework allows for a close-form measure of estimation uncertainty through the error covariance matrix $\Sigma_{\mathbf{e}} = \mathbb{E}[\mathbf{e}\mathbf{e}^T]$, where $\mathbf{e} = \mathbf{m} - \hat{\mathbf{m}}_{ML}$. For the ML estimate, the error covariance is given by

$$\Sigma_{\mathbf{e}} = (\mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A})^{-1} \quad (6.4)$$

where the expected squared error of the i th element of \mathbf{m} is the (i, i) th element of $\Sigma_{\mathbf{e}}$.

6.3 Optimality criteria

In this section, based on the statistical framework of Section 6.2, we consider three compelling choices for the $\text{Cost}(\mathbf{A})$.

1. *Sum of squared errors*: The sum of squared errors (SSE) or $\mathbb{E}[\mathbf{e}\mathbf{e}^T]$ is given by

$$\text{Cost}^1(\mathbf{A}) = \text{Tr}(\Sigma_{\mathbf{e}}) = \sum_i (\Sigma_{\mathbf{e}})_{ii} \quad (6.5)$$

2. *Uniformity of squared errors*: If we wish to design the experiment such that the errors are distributed evenly over all lags, this cost function should be used. It is represented by

$$\text{Cost}^2(\mathbf{A}) = \text{STD}\{(\Sigma_{\mathbf{e}})_{ii}\} \quad (6.6)$$

$$\text{where } \text{STD}\{c_i\} = \sqrt{(1/l) \sum_{i=1}^l \left(c_i - (1/l) \sum_{j=1}^l c_j \right)^2}$$

3. *Condition number of the convolution matrices*: This criterion is different from the other two in which it is merely dependent on the transmitted waveform. The cost function is defined

as the condition number of convolution matrices [Golub and Van Loan, 1989].

$$\text{Cost}^3(\mathbf{A}) = \text{Cond}(\mathbf{A}) \quad (6.7)$$

$$= \frac{\sigma_{\mathbf{A} \text{ (high)}}}{\sigma_{\mathbf{A} \text{ (low)}}} \quad (6.8)$$

where $\sigma_{\mathbf{A} \text{ (high)}}$ and $\sigma_{\mathbf{A} \text{ (low)}}$ represent the highest and lowest singular values of \mathbf{A} , respectively.

6.4 Sequential backward selection algorithm

With the optimality criteria set, we present a modified version of the SBS algorithm which suits our application.

1. Let N be the total number of candidate locations, and n_p be the desired number of short pulses in the code.
2. Let \mathcal{S} to be the set of all N candidate locations.
3. Remove one short pulse at a time. Compute the forward model \mathbf{A} and Σ_ϵ , the error covariance matrix of the measurements. Compute $\text{Cost}(\mathbf{A})$.
4. Let $(S) \leftarrow (S) \setminus \{k^*\} : k^* = \arg \min_{k \in S} \text{Cost}(\mathbf{A}^{(S \setminus \{k\})})$
5. Go to step 3 until n_p short pulses are left in the code.

This approach is suboptimal in the sense that it is not guaranteed to find the combination that minimizes the criterion [Reeves and Heck, 1993]. The reason is that the algorithm has no ability to backtrack. Although no backtracking prunes the search space considerably, this approach presents the problem that once a short pulse is removed from the set, it cannot be added at a later stage. This early commitment to delete a segment overlooks many combinations that might perform better than the selected set. Although the SBS algorithm is suboptimal, it eliminates the combinatoric problem and still provides good results.

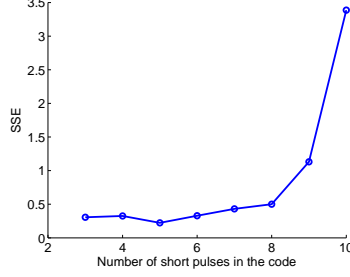


Figure 6.1 Variation of SSE as a function of the number of short pulses present in the code.

6.5 Numerical results

In this section, we present numerical results of the performance of the SBS algorithm. When possible, we compare the performance of SBS with the exhaustive search which always leads to the optimal solution. We assume that the ionosphere is constant throughout all altitudes. Let $T_e = T_i = 1500$ K, and $N_e = 0.5 \times 10^{12} \text{ m}^{-3}$. Let $\text{SNR} = 0.5$, and the pulse length to be $300 \mu\text{s}$.

In the first part of the simulation, we let the lag resolution or the baud length be $30 \mu\text{s}$. Therefore, there will be total of 10 candidate locations for the short pulses in the code. We execute the algorithm described in Section 6.4, where the optimality criterion is chosen to be SSE. Figure 6.1 shows the variation of SSE as short pulses are removed in several iterations. Note that SSE is the highest for the long pulse (iteration 0), and as short pulses are removed, it monotonically decreases until the on-off ratio of 50% is achieved (iteration 5). After that, it slightly increases. Therefore, the optimality criterion reaches its minimum at iteration 5.

Figure 6.2 shows the pulse obtained at each iteration on the left and the number of integrated heights for each time lag on the right. Note that when the on-off ratio reaches below 50%, the number of integrated heights reaches zero for some time lags, which means that the ACF cannot be measured at those lags. This is not a desirable property, as we would like to be able to measure all lags of ACF in an experiment. The lowest on-off ratio that results in nonzero integrated heights for each lag is 50%. Table 6.1 shows the ratio of standard deviation (std) to the average of the number of integrated heights for all time lags at each iteration. According to this table, the optimal pulse that is obtained via the SBS algorithm results in the most uniform distribution of the number of integrated heights for all lags.

In the next part of the simulations, we use the second optimality criterion for the algorithm.

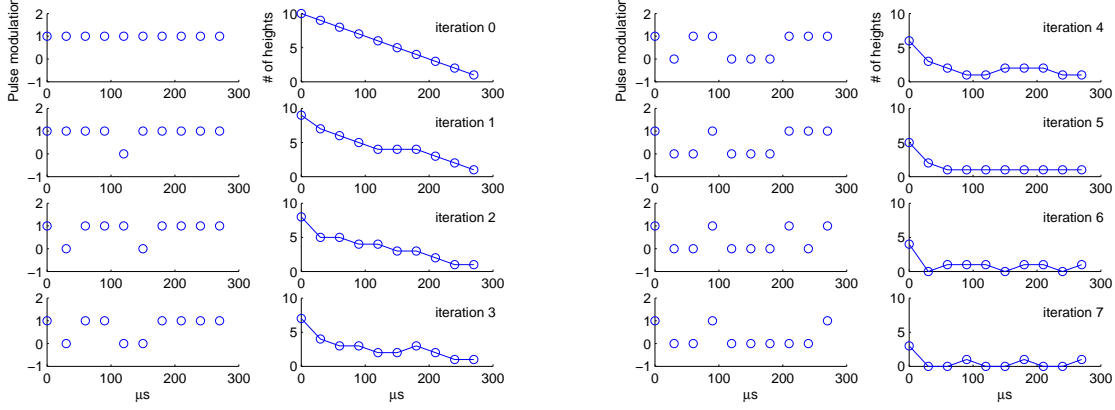


Figure 6.2 Optimal modulated pulse for iterations 0 to 7 is in columns 1 and 3. Number of integrated heights for each time lag is in columns 2 and 4.

Table 6.1 Ratio of standard deviation to the average of the number of integrated heights for all time lags at each iteration. Lag resolution is set to $30 \mu s$, and the length of the pulse is set to $300 \mu s$.

Iteration	0	1	2	3	4	5	6	7
$\frac{\text{std}}{\text{mean}}$ of the # of heights for all lags	0.46	0.47	0.49	0.43	0.42	0.3	0.75	1.5

Figure 6.3(a) shows the variation of the corresponding cost function as a function of iterations. Similarly to the previous case, the minimum happens at the on-off ratio of 50%. Figure 6.3(b) shows the plasma ACF lag error covariance matrix for the optimal modulation.

The change of plasma parameter does not change the results. However, as one expects, the change in the SNR does change the optimal pulse. For instance, Figure 6.4 shows the variation of SSE for $\text{SNR} = 0.01$. When the SNR is low, pulses with very small on-off ratios suffer more from the background noise, and hence the optimal SSE happens at higher on-off ratios. In this case, the minimum happens at iteration 2 when the on-off ratio is 80%.

Next, we consider the pulse length and lag resolution to be $300 \mu s$ and $20 \mu s$, respectively. The variation of the SSE optimality criteria as a function of the number of short pulses in the code is shown in Figure 6.5. Based on this figure, the optimal on-off ratio is 53% in this case. The optimal modulation along with the number of integrating heights for each lag is presented in Figure 6.6(a). Figure 6.6(b) shows the optimal pulse obtained via an exhaustive search. The SSEs for the two

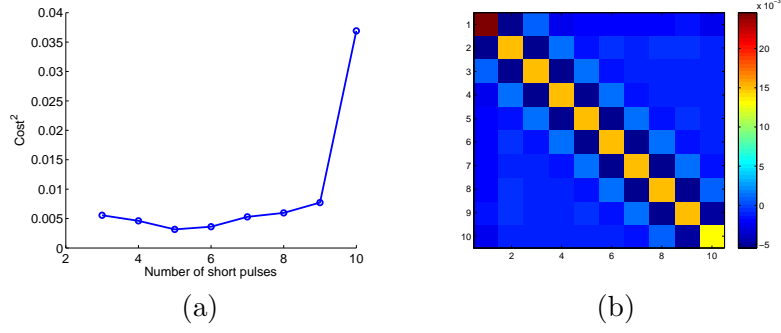


Figure 6.3 (a) Variation of Cost^2 as a function of the number of short pulses in the code. (b) Plasma ACF lag error covariance matrix for the optimal modulation with the criteria Cost^1 and Cost^2 .

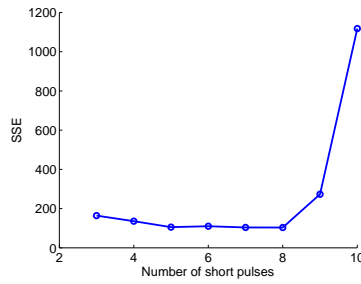


Figure 6.4 Variation of SSE as a function of the number of short pulses in the code for $SNR = 0.01$.

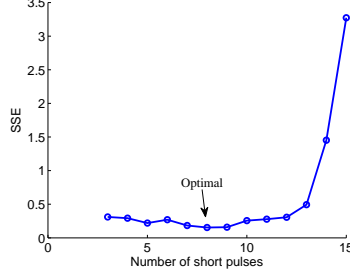


Figure 6.5 Variation of SSE as a function of the number of short pulses in the code for lag resolution $20 \mu\text{s}$.

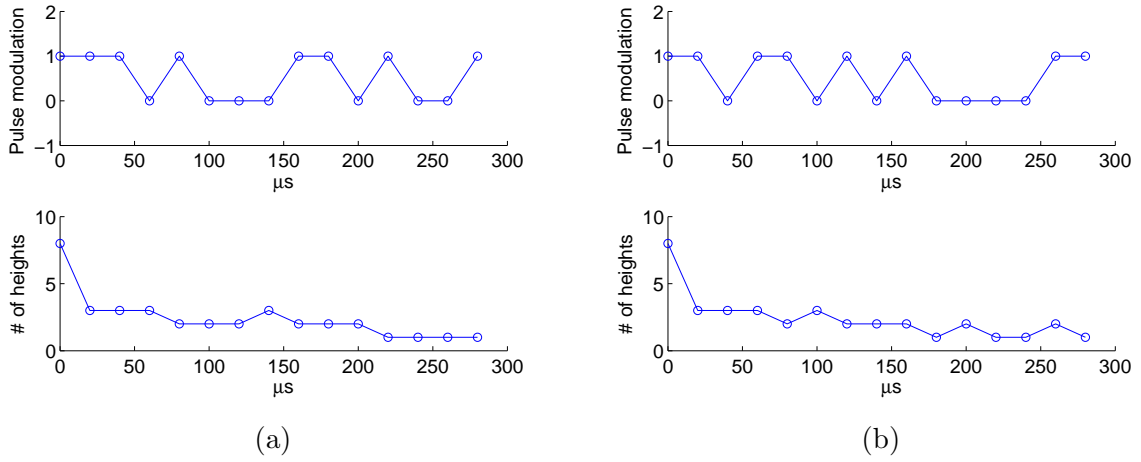


Figure 6.6 (a) SBS selected pulse with lag resolution $20 \mu\text{s}$ (top) and the number of integrated heights for each lag (bottom). (b) Exhaustive search selected pulse with lag resolution $20 \mu\text{s}$ (top) and the number of integrated heights for each lag (bottom).

pulses are very close (1.54 for the SBS, and 1.48 for the exhaustive search), and interestingly the standard deviation to the mean for the number of integrated heights is 0.39 for both.

Finally, in order to provide a comparison with the results of Chapters 4 and 5, we set the pulse length, lag resolution, and on-off ratio to $280 \mu\text{s}$, $10 \mu\text{s}$, and 57%, respectively. For the optimality criterion, this time, we choose the average condition number of convolution matrices for all lags. The difference between this optimality criterion with other criteria is that this criterion is entirely dependent on the waveform rather than the characteristics of the plasma or SNR (although one should note that the on-off ratio of 57% is more appropriate for the medium to high SNR scenarios).

Figure 6.7(a) shows the variation of the mean condition number of convolution matrices for all lags as a function of the number of short pulses in the transmitted waveform. Similarly to SSE, the

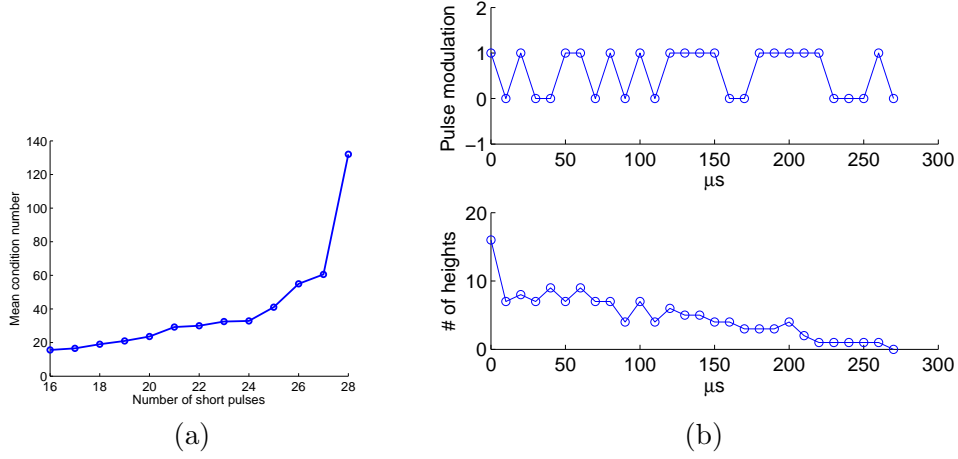


Figure 6.7 (a) Variation of mean condition number of convolution matrices as a function of the number of short pulses in the code for lag resolution $10 \mu\text{s}$. (b) SBS selected pulse with lag resolution $10 \mu\text{s}$ and on-off ratio of 0.57% (top) and the number of integrated heights for each lag (bottom).

mean condition number also decreases monotonically as the algorithm progresses. Figure 6.7(b) shows the optimal pulse obtained from the algorithm along with the number of integrating heights.

In order to compare the performances of this optimally chosen pulse (AM7) and AM3, we present numerical results on parameter estimation. Figure 6.8 shows the estimated parameters when AM7 is utilized in transmission for a high SNR scenario (see Figure 4.6 (a)). From comparing these results with those in Figure 4.12, we realize that AM7 yields a slight improvement over AM3. This slight improvement is due to the slightly lower condition numbers of convolution matrices (see Figure 6.9). The mean condition number is 19 and 16 for AM3 and AM7, respectively. Table 6.2 summarizes the estimation results for the two pulses.

Table 6.2 Estimated errors in physical parameters for a high SNR situation for 50 runs of Monte Carlo simulations for AM3 and AM7.

	Electron temperature (K)	Ion temperature (K)	Electron density (m^{-3})
AM3	30.2	40.1	1.9e10
AM7	26.7	37.2	1.6e10

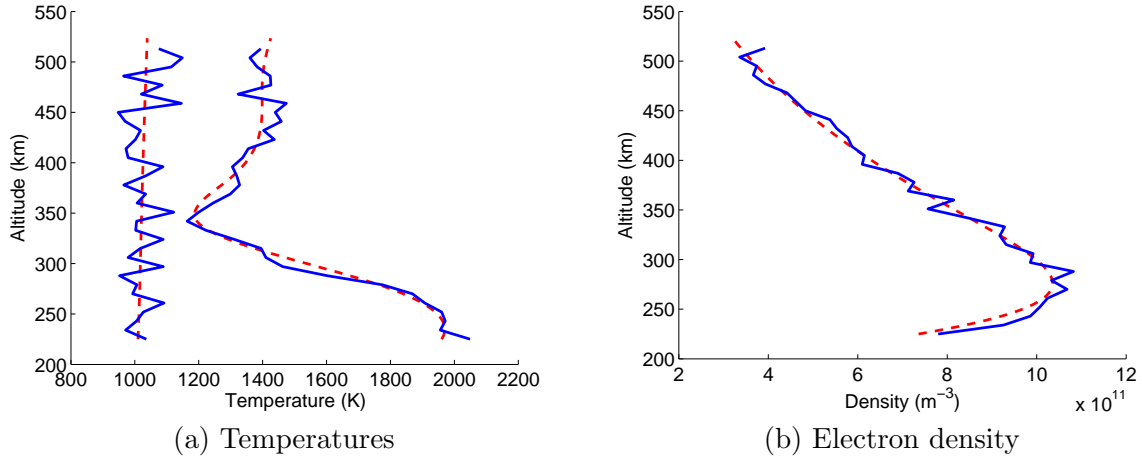


Figure 6.8 Estimated physical parameters (solid lines) for AM7 measurements in a high SNR regime. Original parameters are denoted by dashed lines.

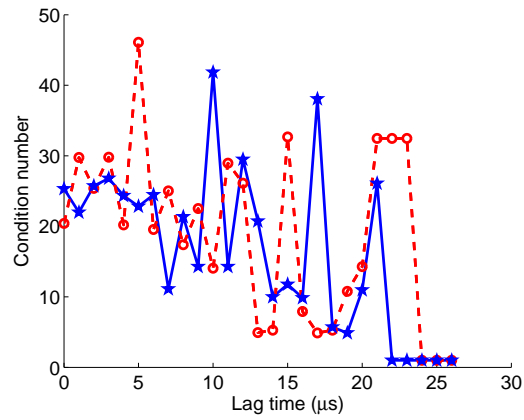


Figure 6.9 Condition number of convolution matrices for AM3 (dashed lines) and AM7 (solid lines) as a function of time lag. The mean condition number is 19 and 16 for AM3 and AM7, respectively.

CHAPTER 7

CONCLUSIONS

This chapter concludes the thesis with a summary of the novel developments and experimental contributions of this work in terms of modeling approach, inversion technique, parameter estimation results, and experiment design.

7.1 Modeling

The first major step of this work was the development of a discrete forward model for F-region incoherent scatter measurements, where long pulses are utilized in transmission. The range smearing imposed on measurements by long pulses was modeled as a 1-dimensional convolution along the range in the simplified case where the receiver sampling is instantaneous. The convolution kernel, i.e. ambiguity function, at each time lag was computed based on the product of the transmitted pulse and its shifted version. It was then discretized in terms of altitude with $\frac{c\Delta\tau}{2}$ increments. The kernels used along with the antenna gain, radiation resistance of the antenna, and the classic electron radius formed the forward model relating the received signal ACF and plasma ACF at individual altitudes.

The forward model was also developed for cases where the receiver sampling was not instantaneous. The model relating the ACF of the received signal to the plasma ACF at several heights and several time delays was computed as a convolution of the products of receiver sampling and transmitted waveform with their shifted versions. The discretization of the forward model was performed in terms of both altitude and time delay.

In Chapter 4, the discretization of the forward model was slightly modified to obtain the resolution supported by the long-pulse measurements. The modified projection matrix was obtained by

averaging of the columns of the original forward model.

7.2 Inversion technique

The next major phase of this research was to develop an efficient hybrid technique that allowed for estimation of plasma parameters by removing the range smearing from measurements. This required overcoming the challenges of the ill-conditioning of the projection matrix associated with long pulses.

The inversion technique incorporated both quadratic and edge-preserving regularization approaches in order to provide smooth ACF lag profiles in the presence of noise while still resolving the sharp gradients. Two regularization parameter selection methods, namely L-curve and UPRE, were studied. The method was validated in the context of simulations, and results demonstrated significant improvements in terms of both accuracy and computational complexity over conventional analysis techniques. In cases where the forward model is modified based on the resolution supported by measurements, no regularization was necessary and the plasma ACF lag profiles were obtained via the least-squares technique.

7.3 Experiment design

Another contribution of this thesis was to develop a technique for optimal modulation design in ISR experiments. The optimal resolution supported by ISR measurements was used as one criterion for the optimal design. The model order selection framework was applied to the problem at hand to find the optimal resolution. The results indicated that, compared with a long pulse, amplitude-modulated codes yielded a finer range resolution with nearly similar parameter estimation errors or smaller estimation errors with the same range resolution. In medium to high SNR scenarios, a smaller on-off ratio of the transmitted waveform was recognized as a determining factor for allowing more freedom in removing range ambiguities as well as resulting in improved statistical accuracy for integrated data in the range and lag directions.

In order to find the optimal amplitude modulation for Arecibo ISR measurements in medium to high SNR scenarios, a modified form of the sequential backward selection algorithm was applied

to the space of all amplitude modulated pulses with a certain on-off ratio. Due to the vast search space, there was no possibility for an exhaustive search. Therefore the problem of finding the optimal amplitude modulation was viewed as an optimization problem. Three optimality criteria, namely, sum of squared errors, uniformity of estimation errors, and condition number of convolution matrices, were considered. The numerical results indicated the effectiveness of the proposed algorithm.

7.4 ISR experiments and parameter estimation results

The final contribution of this work was to implement and conduct several experiments in July 2006, August 2005, and April 2004 using the incoherent scatter radar at Arecibo Observatory, and apply the inversion technique to estimate the plasma parameters. In these experiments, the original mode of MRACF was modified to utilize amplitude modulation. A long pulse and several AM pulses with different on-off ratios were interleaved with a duration of 10 s. Barker codes were also added to the 2006 experiment to provide the ground truth for electron density. Except for the 2004 experiment, the experiments were conducted mostly during the nighttime, where the SNR reached below 0.07 at about 490 km altitude. The data of each coding scheme were analyzed using the optimal resolution achievable in high SNR scenarios. Despite different range resolutions, the mean estimated parameters were nearly the same for measurements from all coding schemes. Moreover, the estimated electron density profiles matched the electron density obtained from the Barker code data. The estimation errors, also, were nearly the same up to 490 km. Above this altitude, the estimation errors increased especially for the AM pulses with smaller on-off ratios. This is expected because the sensitivity of AM measurements to the background noise increases as the SNR decreases. The 2004 experiment was conducted in daytime using only two AM pulses. Because the SNR was much higher during the day, the estimation errors were nearly the same up to about 650 km. The results of all these experiments verified that when the SNR was sufficiently high, compared with an unmodulated long pulse, improved range resolution with nearly the same statistical accuracy was obtained when an amplitude modulation is utilized.

7.5 Future directions

The main goal of this thesis is to accurately and efficiently estimate the key ionospheric state parameters which are required to understand fundamental issues of terrestrial plasma physics such as redistribution of energy and momentum, and coupling within terrestrial upper atmosphere regions. The results of the developed methodology and experiment design of this work can be extended to other ISR facilities to improve the estimation task in other altitude and latitude regions, and to extract many further ionospheric parameters such as electric field strength, conductivity and current, and neutral wind speed. The remainder of this section discusses such extensions.

7.5.1 Arecibo topside mode

Similarly to the MRACF mode, the Arecibo topside mode uses the ion line to measure properties of the ionospheric parameters up to 4000 km in altitude. In the standard configuration, a $500\ \mu\text{s}$ pulse is transmitted every 20 ms, and parameters are measured between 255 and 2600 km [Gonzalez *et al.*, 2004; Isham *et al.*, 2000]. Because of the coarse range resolution, 75 km, at altitudes near 255 km, accurate parameters cannot be estimated as the pulse is too long to resolve the short scale heights at those altitudes.

Utilizing amplitude modulation in topside studies is not favorable because the SNR levels are much lower than those of F-region measurements, and amplitude-modulated waveforms are more sensitive to the noise background level. However, phase modulation can be used in this type of observation. The high correlation between lag estimate errors that typically occurs in F-region measurements is not an issue due to low SNRs. Similar methodology in the optimal experiment design (SBS algorithm) is applicable in this case; the only difference is that the 0's in the AM pulses are replaced with -1's in phase modulation. Derivation of upper bounds for errors and convergence properties of the SBS algorithm under different optimality criteria are challenges ahead. However, once the optimal waveform is chosen, one can form the forward model of the ambiguity and apply the inversion technique developed in this work. The inversion technique can also be easily extended to include different temperatures and velocities for oxygen and hydrogen ions.

7.5.2 Advanced modular incoherent scatter radar

Among all observation facilities, the first advanced modular incoherent scatter radar (AMISR), the Poker Flat incoherent scatter radar (PFISR), has been recognized as one of the most valuable ground-based diagnostic tools for studying the Earth's polar upper atmosphere within the geoscience community (e.g. [Vadas and Nicolls, 2008; Nicolls and Heinselman, 2007]). The reason is the unique steering and beam-forming capabilities of the AMISR technology. Located at the Poker Flat Research Range (65.13°N, 147.47°W) near Fairbanks, Alaska, and operating at 450 MHz, PFISR is a state-of-the-art phased array radar which allows the experimenter to steer the beam on a pulse-to-pulse basis using phased-array techniques, providing a powerful extension over typical ISR approaches, which only offer high-resolution measurements in one direction or reduced time resolution in multiple directions through physical steering. The PFISR gives the scientific community a great opportunity to advance our understanding of the polar ionosphere. Fully exploiting the capabilities of this remarkable observation facility, however, requires advances in radar measurement analysis techniques.

One extension of the present work is to estimate latitudinal and temporal variation of electric fields in the F- or E-regions of the polar ionosphere with the final goal of producing estimates of the convection field as a function of magnetic latitude [Holt and van Eyken, 2000; Foster et al., 1981]. For this purpose, one needs to obtain line-of-sight (LOS) velocities from multiple directions [Heinselman and Nicolls, 2008]. The LOS velocities can be estimated from a fit to the full incoherent scatter auto-correlation function (including both the real and imaginary parts). Once LOS drifts are obtained, vector drifts must be estimated using multiple spaced beams. Currently, long-pulse experiments are conducted with PFISR to make these types of observations. However, with a combination of amplitude modulation experiments and edge-preserving inversion techniques, accurate, efficient, and high-resolution parameter estimation will become possible. With the efficiency of the proposed work, one can perform the estimation task in real time, which is very important especially in high latitudes, where the ionosphere is subject to constant change.

7.5.3 Jicamarca radar

Since it began operating in 1961, the Jicamarca Radio Observatory has been an important source of plasma density, temperature, composition, and drift measurements in the equatorial zone. Its distinction from other ISR facilities is due partly to its proximity to the magnetic dip equator. The Jicamarca radar consists of a modular, phased-array antenna with a main lobe that can be directed perpendicular to the geomagnetic field [Hysell, 2000]. The narrowness of the incoherent scatter spectrum for perpendicular backscatter permits very accurate measurements of cross-field plasma drift [Kudeki *et al.*, 1999]. Because of the operating frequency of 50 MHz, the auto-correlation time of the incoherent scatter signal when observed a few tenths of a degree or more off the perpendicular is of the order of 1 to 2 ms [Hysell, 2000]. This implies that the range resolution of a conventional long-pulse experiment is about 300 km for the shortest lags.

Because the correlation time of the ionosphere at Jicamarca is completely different from that of the Arecibo radar, the requirements of F-region observation such as the length of the transmitted waveform and lag resolution will be completely different in the two cases. Nevertheless, our developed inversion technique and modulation design is still applicable to the Jicamarca radar for plasma temperature, density, composition, and LOS velocity measurements, especially in situations when the ionosphere is changing rapidly and long integration time, which is required for alternating code experiments, is not possible.

REFERENCES

- Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, in *Proc. 2nd Inter. Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki, pp. 267–281, Akademiai Kiado, Budapest.
- Bittencourt, J. A. (2004), *Fundamentals of Plasma Physics*, Springer, New York.
- Blahut, R. (2004), *Theory of Remote Image Formation*, Cambridge University Press, Cambridge, UK.
- Burnham, K., and D. Anderson (1998), *Model Selection and Inference - A Practical Information Theoretic Approach*, Springer, New York.
- Coleman, T. F., and Y. Li (1996), An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optimization*, 6(2), 418.
- Couivreux, C., and Y. Bresler (2000), On the optimality of the backward greedy algorithm for the subset selection problem, *SIAM J. Matrix Anal. Appl.*, 21(3), 797–808.
- Cover, T. M., and J. A. Thomas (1991), *Elements of Information Theory*, John Wiley and Sons, New York.
- Demoment, G. (1989), Image reconstruction and restoration: Overview of common estimation structures and problems, *IEEE Trans. Acoustics, Speech, and Signal Process.*, 37(12), 2024–2036.
- Dougherty, J., and D. Farely (1960), A theory of incoherent scattering of radio wave’s by a plasma, *Proc. R. Soc. London Ser A*, 259, 79.
- Dougherty, J., and D. Farely (1961), A theory of incoherent scattering of radio waves by a plasma, II, *Proc. R. Soc. London Ser A*, 263, 238.
- Farley, D. (1966), A theory of incoherent scattering of radio waves by a plasma, 4, *J. Geophysical Research*, 71, 4091.
- Farley, D. (1969), Incoherent scatter correlation function measurements, *Radio Sci.*, 4, 935.
- Farley, D. (1972), Multi-pulse incoherent scatter correlation function measurements, *Radio Sci.*, 7(6), 661.
- Foster, J. C., J. R. Doupnik, and G. Stiles (1981), Large scale patterns of auroral ionospheric convection observed with the Chatanika radar, *J. Geophysical Research*, 86, 11,357–11,371.
- Golub, G., and C. F. Van Loan (1989), *Matrix Computations*, Johns Hopkins U. Press, Baltimore, MD.

- Gonzalez, S. A., M. P. Sulzer, M. J. Nicolls, and R. B. Kerr (2004), Solar cycle variability of nighttime topside helium ion concentrations over Arecibo, *J. Geophysical Research*, *109*, A07,302.
- Hagfors, T. (1961), Density fluctuations in a plasma in a magnetic field, with applications to the ionosphere, *J. Geophys. Res.*, *66*, 1699.
- Hagfors, T. (1971), A theory of collision dominated electron density fluctuations in a plasma with applications to incoherent scattering, *Phys. Fluids*, *14*, 1143.
- Harwit, M., and N. Sloane (1979), *Hadamard Transform Optics*, Academic Press, New York.
- Heinselman, C. J., and M. J. Nicolls (2008), A Bayesian approach to electric field and E-region neutral wind estimation with the Poker Flat Advanced Modular Incoherent Scatter Radar, *Radio Sci.*, *43*, RS5013.
- Holt, J., D. Rhoda, D. Tenebaum, and A. van Eyken (1992), Optimal analysis of incoherent scatter radar data, *Radio Sci.*, *27*(3), 435.
- Holt, J. M., and A. P. van Eyken (2000), Plasma convection at high latitudes using the EISCAT VHF and ESR incoherent scatter radars, *Annales Geophysicae*, *18*(9), 1088–1096.
- Huuskonen, A., and M. Lehtinen (1996), The accuracy of incoherent scatter measurements: Error estimates valid for high signal levels, *J. Atmosphere and Terrestrial Phys.*, *58*, 453.
- Hysell, D. L. (2000), Incoherent scatter experiments at Jicamarca using alternating codes, *Radio Sci.*, *35*, 1425–1436.
- Isham, B., C. A. Tepley, M. P. Sulzer, Q. H. Zhou, M. Kelley, J. Friedman, and S. Gonzalez (2000), Upper atmospheric observations at the Arecibo Observatory: Examples obtained using new capabilities, *J. Geophysical Research*, *105*, 609.
- Kamalabadi, F., W. C. Karl, J. L. Semeter, D. M. Cotton, T. A. Cook, and S. Chakrabarti (1999), A statistical framework for space-based EUV ionospheric tomography, *Radio Sci.*, *34*, 437–448.
- Karl, W. (2000), Regularization in image restoration and reconstruction, in *Handbook of Image and Video Processing*, edited by A. C. Bovik, pp. 141–160, Academic Press, Inc., Orlando, FL.
- Kendall, M., and A. Stuart (1979), *The Advanced Theory of Statistics*, vol. 2, Griffin, London.
- Konishi, S., and G. Kitagawa (2008), *Information Criteria and Statistical Modeling*, Springer, New York.
- Kudeki, E. (2003), ECE 458 lecture notes, application to radio wave propagation, University of Illinois at Urbana-Champaign, Urbana, IL.
- Kudeki, E., S. Bhattacharyya, and R. Woodman (1999), A new approach in incoherent scatter F-region $E \times B$ drift measurements at Jicamarca, *J. Geophysical Research*, *104*, 28,145.
- Lagendijk, R. L., and J. Biemond (2000), Basic methods for image restoration and identification, in *Handbook of Image and Video Processing*, edited by A. C. Bovik, pp. 125–139, Academic Press, Inc., Orlando, FL.
- Lehtinen, M., and A. Haggstrom (1987), A new modulation principal for incoherent scatter measurements, *Radio Science*, *22*, 625.

- Lehtinen, M., and A. Huuskonen (1986), The use of multipulse zero lag data to improve incoherent scatter radar power profile accuracy, *J. Atmospheric and Terrestrial Phys.*, 48, 787.
- Lehtinen, M., and A. Huuskonen (1996), General incoherent scatter analysis and GUIDAP, *J. Atmospheric and Terrestrial Phys.*, 58, 435.
- Lehtinen, M., A. Huuskonen, and J. Pirttilä (1996), First experiences of full-profile analysis with GUIDAP, *Annales Geophysics*, 14(12), 1487.
- Levenberg, K. (1944), Method for the solution of certain problems in least squares, *Quart. Appl. Math.*, 2, 164.
- Marill, T., and D. M. Green (1963), On the effectiveness of receptors in recognition systems, *IEEE Trans. Information Theory*, 9(1), 11–17.
- Marquardt, D. (1963), Algorithm for least squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, 11, 431.
- Miller, A. (2002), *Subset Selection in Regression*, Chapman & Hall/CRC, FL.
- Nicolls, M. J., and C. J. Heinselman (2007), Three-dimensional measurements of traveling ionospheric disturbances with the Poker Flat incoherent scatter radar, *Geophys. Res. Lett.*, 34, L21,104.
- Nikoukar, R., F. Kamalabadi, E. Kudeki, and M. Sulzer (2008), An efficient near-optimal approach to incoherent scatter radar parameter estimation, *Radio Sci.*, 43, RS5007.
- Nygren, T. (1996), *Introduction to Incoherent Scatter Measurements*, Inverse Publications, Sodankylä, Finland.
- Papoulis, A. (1986), *Probability, Random Variables and Stochastic Processes*, McGraw Hill Book Company, Inc., New York.
- Poor, V. (1994), *An Introduction to Signal Detection and Estimation*, Springer, New York.
- Reeves, S., and L. Heck (1993), Selection of observations in signal reconstruction, *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 3, 444–447.
- Reeves, S. J., and Z. Zhe (1999), Sequential algorithms for observation selection, *IEEE Trans. on Signal Processing*, 47(1), 123–132.
- Sharif, B. (2004), A hybrid technique, fundamental limitations, and optimal design, M.S. thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.
- Sharif, B., and F. Kamalabadi (2005), Optimal discretization resolution in algebraic image reconstruction, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 803, 199.
- Sulzer, M. (1986), A phase modulation technique for a sevenfold statistical improvement in incoherent scatter data-taking, *Radio Sci.*, 21(4), 737.
- Vadas, S. L., and M. J. Nicolls (2008), Using PFISR measurements and gravity wave dissipative theory to determine the neutral, background thermospheric winds, *Geophys. Res. Lett.*, 35, L02,105.

- Vallinkoski, M. (1988), Statistics of incoherent scatter multiparameter fits, *J. Atmospheric and Terrestrial Phys.*, 50, 839.
- Vogel, C. R. (2002), *Computational Methods for Inverse Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Vogel, C. R., and M. E. Oman (1996), Iterative methods for total variation denoising, *SIAM J. Sci. Comput.*, 17, 227–238.
- Woodman, R. F. (1991), A general statistical instrument theory of atmospheric and ionospheric radars, *J. Geophys. Res.*, 96, 7911.